DOCUMENT RESUME

ED 174 206 IR 007 480

TITLE	Integrated Library Network System for the Commonwealth of Virginia--A Feasibility Study.
INSTITUTION	Virginia State Library, Richmond.
NOTE	310p.
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	Computers; *Electronic Data Processing; Federal State Relationship; Flow Charts; Library Automation; Library Cooperation; *Library Networks; Maps; *Statewide Planning
IDENTIFIERS	Turnkeying; *Virginia

ABSTRACT

This feasibility study of an integrated statewide library network system consists of two main parts: the body of the report, which includes an introduction to the background and method of the study, the findings, and the modular plan for implementation; and supplementary materials presented in seven appendices reviewing basic facts. These appendices cover library applications of automation, bibliographic services, network components and standardization (both hardware and software), data management software, hardware, and turn-key systems; the last appendix is a glossary of terms. Also provided is a lengthy addendum to the report that presents the requirements and recommendations for such a network--i.e., the data processing requirements for each component within a network with hardware and software requirements for the component and time and cost projections, and a modular approach recommended for the eventual development of a fully integrated library network system. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174206

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

INTEGRATED LIBRARY NETWORK SYSTEM
FOR
THE COMMONWEALTH OF VIRGINIA

-- A FEASIBILITY STUDY --

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dorothy Fuller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR007480

TABLE OF CONTENTS

INTEGRATED LIBRARY NETWORK SYSTEM

I. INTRODUCTION

. Charge of Study 1.1
. Goals of Action 1.2
. Federal-State Relationships 1.4
. Advantages of Cooperation 1.7
. Cooperation in Virginia 1.8
. Union Lists 1.12
. Input to the Study 1.14

II. FINDINGS AND CONCLUSIONS

. Prefatory Note 2.1

. Findings

.. Usefulness of a Network 2.2
.. Bibliographic Data Base 2.2
.. Procedures Toward Development 2.5

. Conclusions 2.8

III. A MODULAR APPROACH TO A NETWORK

. Introduction 3.1
. Module I 3.4

.. Growth Union List 3.6
.. Conversion Techniques 3.12
.. Data Base Manipulation 3.14
.. Library Automation 3.16

. Module II 3.27
. Module III 3.32

IV. APPENDICIES

A. Applications

. Acquisitions A.1
. Cataloging A.7

- Circulation A.17
- Reference A.21
- Inter-Library Loan A.24

B. Bibliographic Services

- History B.1
- Search Techniques B.2
- Timing Factors B.7
- Cost B.8
- Products B.9
- Future Services B.10

C. Network Components and Standardization

- Introduction C.1
- Hardware Components C.4
  - Circuits C.4
  - Carriers C.5
  - Interactive Terminals C.6
  - Information Transfer Rate C.7
  - Analog/Digital Adapters C.7
  - Function of a Modem C.7
  - Modulation Techniques C.8
  - Asynchronous Modems C.8
  - Synchronous Modems C.8
  - Protocols C.9
  - Conditioning/Equalization C.9
  - Simplex/Duplex C.10
  - Multiplexing C.11
  - Polling C.11
  - Network Processors C.13
  - Front End Processors C.13
  - Concentrators C.14
  - Switching Processors C.15
  - Error Detection and Treatment C.16
- Software Components C.17
  - Minicomputers C.17
  - Front End Processors C.18
  - Remote Concentrators C.18
  - Communication Processors C.19
  - Communication Control Procedures C.19
  - Phase of Control C.20

.. Error Control C.20
.. Common Headings C.21
.. Protocols C.22

. Standardization C.23
. Conclusions C.30

D. Data Base Management Software

. Definition D.1
. General Guidelines D.4
. ADABAS D.6

.. Characteristics D.7
.. Hardware/Software Requirements D.8
.. Pricing D.8
.. Users D.8

. Total D.9

.. Characteristics D.10
.. Hardware/Software Requirements D.11
.. Pricing D.11
.. Users D.11

. IMS D.12

.. Characteristics D.13
.. Hardware/Software Requirements D.14
.. Pricing D.14
.. Users D.14

. Conclusions D.15

E. Hardware

. Overview E.1
. Evaluation Tools E.4

F. Turn-Key Systems

- Overview F.1
- Data Phase F.2
- Cincinnati Electronics F.2
- Systems Control, Inc. F.3
- DECICOM F.3
- GAYLORD F.3
- ICC/PLESSEY F.4
- CLSI F.4
- Universal Library Systems, Inc. F.4
- Innovated Systems, Inc. F.5

G. Glossary of Terms G.1

FOREWORD

The following report is organized to allow the reader access first to the matters of major interest: the essential findings of the study and the suggested plan of implementation. It is accordingly divided into two main sections. The first constitutes the body of the report: after an introduction to the background and method of the study, the findings and plan for implementation are set out. Materials supplementary to the considerations in the body of the report are appended in the second part. These appendices review for the reader the basic facts and elements of matters significant in the first part: data base management systems, communications networks, host processors, turn-key systems, bibliographic data base facilities, and library applications of automation. The list of appendices is given in the table of contents.

INTRODUCTION

The charge of this study is to report on the feasibility of establishing a statewide library network for the Commonwealth of Virginia. The network in question is defined as an organization capable of serving all types of libraries and concerned with providing its members products and services derived from computer and telecommunications technology. This technology would serve to maintain a Virginia bibliographic data base (or bases) organized for a variety of library applications at the state, regional, and local levels. Findings regarding the feasibility of such a network are given at the end of this section along with a proposed guideline to achieve a network. Prior to that, however, we will review the national and state background as a preface to understanding the interests, limitations, and particular methods of the study.

Library networks are simply developments in the larger and older process of library cooperation. In the 20th century, library cooperation has been accelerated and facilitated at all levels far beyond anything achieved previously. This is in part due to changes in society which have placed extremely heavy demands on the individual library and library system and to changes in technology which allow efficient and rapid manipulation and communication of information. The extension of education through all age groups and levels of society, the explosion in the amount of recorded information and in the need to obtain it on the part of the public for personal and professional reasons, the continuing inadequacy of library budgets, and the inflationary costs of materials and personnel have indicated with unavoidable clarity that no individual library or library system has the resources to meet the increasing and increasingly varied demand. The only appropriate response has been to combine efforts and share resources.

At the national level, the concerns which lie behind this movement and the direction it seeks to take have been expressed by the National Commission of Library and Information Science. In 1975, the Commission published its program objectives in a report entitled Toward a National Program for Library and Information Service: Goals for Action. This document has since become something of a guiding text for people interested in library networking at all levels. In its report, the Commission sets out the basic assumptions for the sharing of information resources and the establishment of a national network:

> First, that the total library and information resources in the United States is a national resource which should be strengthened, organized and made available to the maximum degree possible in the public interest. This national resource is the cumulated and growing record of much of our nation's and, indeed, the world's total cultural experience--intellectual, social, technological, and spiritual.
>
> Second, that all people of the United States have the right, according to their individual needs, to realistic and convenient access to this national resource for their personal enrichment and achievement, and thereby for the progress of society.

> Third, that with the help of new technology and with national resolve, the disparate and discrete collections of recorded information in the United States can become, in due course, an integrated nationwide network.
>
> Fourth, that the rights and interests of authors, publishers, and other providers of information be recognized in the national program in ways that maintain their economic and competitive viability.
>
> Fifth, that legislation devised for the coherent development of library and information services will not undermine constitutionally-protected rights of personal privacy and intellectual freedom, and will preserve local, state, and regional autonomy.

In consonance with these assumptions the Commission has developed two major program objectives: (1) to strengthen or create, where needed, the human and material resources that are supportive of high quality library and information services; and (2) to join together the library and information facilities in the country, through a common pattern of organization, uniform standards, and shared communications, to form a nationwide network.

It sees the following as indicative of the need for greater cooperation:

1. The growth of libraries in the United States has been fragmented and uneven, leading to waste and duplication of the National knowledge resource and, for lack of common standards, creating obstacles to a cohesive national system.
2. The distribution of library services is correlated with that of population and financial support. While some people have easy access to rich resources, others still lack the most elementary forms of service.

3. The problems of people who lack even the most basic information services or are served only marginally must be identified and addressed.
4. There is a limit to self-sufficiency in the ability of any library, even the largest public or research library, to satisfy its constituents.
5. Special libraries with work-related goals serve at present only limited clienteles.
6. Greater collaboration should be developed among libraries and the commercial and other private sector distributors of the newer information services.
7. Funding at every level is inadequate. A major change in Federal policy is needed to ensure mutually reinforcing funding formulas.
8. New Federal legislation should give local libraries the incentive to join larger systems outside of their immediate jurisdictions.

In proposing a nationwide networking effort, it provides a schematic of federal and state relationships pertinent to various functions:

FUNCTION	FEDERAL	STATE
Goal	To foster cooperative programs within each state and among states by providing federal funding to inaugurate or continue intertype networks at appropriate levels, to demonstrate and test and evaluate them, to support them with technology and to help them move toward financial self-sufficiency.	To develop and operate networks of library and information resources within the state with federal funds expressly earmarked for this purpose within the first three years of the law, and provide for interfacing the intrastate networks with multistate and national networks of all types.

FUNCTION	FEDERAL	STATE
Funding	To increase funding and aid in the operation and improvement of basic library and information services at all levels.	To engage in funding programs which ensure that local requirements for materials and information services sufficient to meet the needs of their constituencies are indeed met.
Planning and Coordination	To coordinate the Federal Program with the states and with multistate groups.	To establish and maintain mechanisms for coordinating state programs with Federal and multistate Programs.
Network Development	To create a Federal Library and information network that will make available the information resources of federal agencies and to coordinate and develop access mechanisms to the resources of the nation.	To create a statewide library and information network that will make available the information resources of state government agencies to the local level and to coordinate and develop access mechanisms to the resources of a region.
Bibliographic Service	To establish and/or support those bibliographic services of nationwide concern that are best carried out on a centralized basis.	To provide means by which national bibliographic services may be used within the state or among states.
Unique Collections	To designate and help sustain unique, national collections and access mechanisms that all states may use as back-up to their own resources.	To designate and sustain for service to the people of the state major collections and access mechanisms for wider internal use and greater self-sufficiency within the state.

FUNCTION	FEDERAL	STATE
Standards	To establish and promulgate technical standards and to facilitate inter-connection across state lines.	To develop means for erasing artificial barriers to cooperation. To support and monitor observance and implementation of standards set by the national Program.
Computer Support	To help each state or multistate region establish computer support services for technical processing and public service purposes. To provide electronic access to the national data base.	To produce and/or provide access to union lists, cumulative indices, directories, and other bibliographic tools to support rapid identification, access, and referral within the state. To coordinate and assist in developing computer support for individual library processing.
Telecommunications Support	To provide telecommunications at low tariff rates, for interstate library and information exchange.	To foster and facilitate low cost communications between and among libraries and other information facilities within the state.
Research and Development	To perform and/or support research, development, and evaluation in areas of nationwide concern or potential; to monitor developments in library and information services; and to disseminate information relating to the above activities.	To perform and/or support research, development, and evaluation in areas of statewide or multistate concern or potential; to monitor developments in library and information services; and to disseminate information relating to the above activities.

FUNCTION	FEDERAL	STATE
Education and Training	To fund education and training programs in modern library and information sciences through fellowships, intern programs, etc.	To provide guidance to federal and state authorities in the development of educational standards, the definition of needed skills, and the installation and evaluation of feed-back mechanisms. States alone or in cooperation with other states have a responsibility for continuing education.
Publicity and Promotion	To plan and carry out programs to inform potential users of the services available through the national network and to actively encourage the use of these services.	To plan and carry out educational and promotional programs tailored to the state and region.
Evaluation	To continually assess the progress of the National Program and to change plans and directions accordingly.	To monitor and evaluate the effectiveness of state and regional programs vis a vis the National Program and plan new state programs where needed.

According to the National Commission, some of the advantages which would accrue to a state participating in this effort are:

1. More information for its residents than it could possibly afford to amass through its own capital investment
2. Reduced interstate telecommunication costs
3. Access to computer software, data bases and technical equipment
4. Compatibility with national programs

5. Matching funding for bringing state and local resources up to acceptable standards
6. Matching funding to initiate network operations
7. The ability to invest mainly in immediate state and local needs while relying upon the national network for specialized material and services.

Neither states nor regions had, or course, waited for the National Commission before embarking on their own systems. In some states, very elaborate inter-library loan and reference networks had been functioning for years (e.g., NYSILL in New York and ILLINET in Illinois). State networks have been interested in the full range of library needs to be met by cooperative action. To the extent that some of these needs can be addressed by teleprocessing, state networks have become involved with arranging for computer and telecommunications services and products. For the most part, this has meant brokering the services of outside network facilities like OCLC for on-line applications (the exception here is the State of Washington which has established its own data base facility) and obtaining machine readable records of the holdings within the state for the production of union lists or catalogs. The list of such networks (and by state network we here mean a distinct organization specifically designated as the center for statewide network activity) is already long and continues to grow. Virginia's interest in networking is, therefore, no isolated phenomenon, but is part of a general movement apparent nationally.

In the case of Virginia, the history of library cooperation exhibits the same evolutionary trend toward networking seen elsewhere in the country. Groups within the state have been formed to address the concerns of either a single type of library or of many types (academic, public, school, special). There are, for example, six consortia presently established for cooperative support of continuing higher education programs in the state. These are primarily composed of academic libraries, although in some instances an effort has been made to include public libraries. Public library directors in the Richmond area and in the Tidewater area have formed their own groups for discussion and action regarding mutual needs. In the Tidewater area, a study has been undertaken for the establishment of a regional network which would include both academic and public libraries. For systems near Washington, D.C. libraries of all types participate in the cooperative programs of the Metropolitan Washington Council of Governments. The Library Advisory Committee of the State Council of Higher Education has brought together the library directors of all state 4-year institutions of higher education and selected representatives from the community college system and private 4-year colleges. Their goals for cooperation include: state-wide collection development, state-wide access to serials holdings, inter-library loan, central depositories, and resource facilities. The State Library has taken responsibility for several significant developments in library cooperation.

It now supports the Virginia Medical Information Service which makes medical information available to all health practitioners in the state. Reference questions and requests for material are serviced by two resource facilities: the University of Virginia Medical Center and the Medical College of Virginia at VCU. These centers have the capability of computer searching of medical literature. In 1977, the State Library established a state-wide inter-library loan referral center at the General Library Branch. This service operates for libraries of all types using TWX and WATS for long distance communication.

Virginia libraries have also participated in several inter-state cooperatives: the Cooperative College Library Center, Afro-American Materials Project, Committee on Libraries of the Southern College University Union, Committee of Libraries of the Washington Technological Consortium, Libraries Technical Committee of the Metropolitan Washington Council of Governments. Twenty-five libraries are members of the Southeastern Library Network, thereby participating in the more than 1500 member network facility at OCLC.

As even this brief review should indicate, Virginia libraries have engaged in a considerable amount of cooperation. Most of these cooperatives, however, have been regional or, if state-wide, have been primarily concerned with libraries of a single type. Except for the inter-library loan center at the State Library, no organization exists with the authority and funding to plan for and implement a coherent state-wide system of service for libraries of all types.

The lack of such an organization and the need for it is very evident in the area of library automation. The automation of library operations provides a natural occasion for cooperation both in order to share the effort and expense required to develop and implement systems and in order to take advantage of the unique ability of the computer to rapidly process and make generally available information received from a great number of different locations and sources. In spite of this, the automated systems presently functioning in the state have, except for the shared cataloging system through SOLINET, been developed in isolation. Such random and uncoordinated activity is not unique to Virginia--it is recognized everywhere as a problem requiring serious attention.

Recently, efforts have been made in Virginia to address this problem. As noted earlier, libraries in the Tidewater area have expressed a serious interest in developing a network for public and academic libraries of the region. Their concern is to establish a common data base capable of serving circulation, inter-library loan, cataloging, and acquisition needs. They are concerned also that such a network be able to serve as a component in any state-wide system. Libraries in the western and northern section and in the capital area of the state have also been exploring the possibilities for common systems. The

State Library has been active too. In February of this year, it conducted a state-wide survey regarding present resources and attitudes toward library automation and cooperative services. As a step toward networking, it has arranged for the regular receipt of archival tapes from the SOLINET libraries in Virginia. These tapes will be merged and a union catalog in COM form produced for distribution to libraries throughout the state. The merging of these tapes constitutes the beginning of a bibliographic data base for Virginia. In addition to its action regarding the SOLINET tapes, the state has commissioned this study to investigate the prospect for further state network development. The specific concerns and procedure of the study are set out in the part which follows.

In its statement to the Department of Management Analysis and System Development, explaining the need for a study of networking, the State Library notes: "Libraries already automated and those preparing to automate are strongly urging the State Library to develop guidelines and set standards for hardware selection, software development, computer protocol and language, bibliographic description, and means of access. These libraries are keenly aware of the need for compatability and access to other data bases. They do not wish to be isolated or bypassed in the eventual and inevitable development of networks, nor do they wish to expend funds in development of data bases and systems that can more wisely be stretched in cooperative one-time effort." It concludes that what is needed is "a carefully developed plan for the design, structure, implementation, and support of a state-wide network so that libraries might share the knowledge and expertise already existing in the state."

As part of this effort, the present study is commissioned to investigate the computer and telecommunications considerations pertinent to a feasible library network design. As such, the study examines the technical components of networking: databases, database management systems, hardware components, supporting software, and telecommunications.

These technical concerns operate, of course, within the context of the library community and its needs, these are adverted to throughout the study. But answers to many of the questions surrounding library policy and practice as well as detailed determinations regarding costs and methods of implementation must wait upon further action at the state level. At present, the network is only projected: there are no actual members, nor any group constituted to speak for a potential membership, nor any formal structure which could bring together such a group. The process by which further decisions will be made is unclear. Unclear also are the limits and basis for funding. The study, therefore, focuses on the technical base for a network, deriving a plan capable of accommodating a variety of library applications in an orderly manner, even if the actual specifications remain to be settled. The appropriateness of this strategy has been further underlined by clarification from the State Library subsequent to the original contract for this study, indicating that questions specific to the constitution and governance of the membership were beyond the scope of the study.

In reaching its conclusions, the study has proceeded by examining the present state of library automation in Virginia and the current and prospective state of the art in library networking. This has involved a review of the literature, of various papers, reports, and documents, and discussion with those engaged in activities significant for this study. Letters requesting information pertinent to library automation and network activities were sent to 15 multi-state networks, 21 states having established networks or inter-library cooperative programs, 18 out-of-state libraries having significant library automation activities, and 30 vendors of library automation services. The directors of the Indiana Library Network and the Michigan Library Consortium were contacted by phone to discuss the Virginia project and to request information regarding their activities. Appropriate people at the Library of Congress' Network Development Office, at the National Commission for Library and Information Science, at the Council on Library Resources, and at the Metropolitan Council of Governments were also contacted by phone to request information. Research was carried out at the Library Services Branch of the U.S. Office of Education (which administers LSCA funds) and at the Library of the School of Library and Information Science at the University of Maryland in order to supplement material available from the State Library and from the ERIC file at Virginia Commonwealth University.

The American Library Association meetings in Chicago were attended in order to gather information about recent developments. The Network conference of the IEEE in Washington, D. C. was also attended to investigate the state of the art in teleprocessing networks more generally. Visits were made to investigate the systems at the Washington Library Network, at Stanford's BALLOTS, and at OCLC.

Meetings were held at various places in Virginia in order to determine the present state of library automation and to see a representative sample of systems in operation. Visits were made to: Virginia Polytechnic Institute and State University, the University of Virginia, Old Dominion University, Northern Virginia Community College, Arlington Public Library, Fairfax Public Library, Fairfax School Media Center, Henrico Public Library, Eastern Shore Public Library, Norfolk Public Library, Tidewater Area directors, the Philip Morris Research Library, and the State Council of Higher Education in Richmond.

In collecting information, we have been primarily interested in discovering the most useful and cost-effective approach to a state-wide network. To this end, we have investigated existing machine readable records of Virginia holdings and the prospect of merging them into a Virginia data base, other data bases through which Virginia's could be updated, management systems available to organize the data base for use, the telecommunications required to access the Virginia data base, network applications, given a Virginia data base managed for state-wide access, turnkey systems able to serve as components for network applications, and the schedule of implementation, given the foregoing considerations.

The section which immediately follows in the report sets out the findings of the study and the suggested plan for implementation. Before turning to it, however, the reader's attention is called to the section on applications which is provided as a review of the uses and potential applications of library automation and automated networks. This review is placed as the first appendix immediately following the body of this report.

FINDINGS

Prefatory Note:

The finding from this study which has had the greatest impact on one of its proposed objectives (i.e., to define network configuration and costs) is that the preliminary work required for such an objective had not been done. Those conducting the study had a choice to make:

1. Produce a group of designs, placing host processors, communication processors, terminals, line connections; compute the hardware costs, line costs; pre-define the applications that could be run on such a configuration and compute software costs. For such designs and projections to have any real value, information is required of a kind not presently available. The survey made by the State Library interests and involvement in networking and library automation in the state was not adequate for the purpose. Further, no group able to speak for the potential membership existed, and the information gathered from individuals represented only a scattering of individual opinions. The alternative was:
2. Set out a realistic plan of action, given the present capabilities of the state and the potential uses of the network, and indicate the specifications required for network design.

We chose the second, deciding that it would yield results more realistic and, therefore, more valuable. It addresses more effectively the ultimate criterion for the study - feasibility. A fuller explanation is given in Section III of the findings.

FINDINGS

I. The Usefulness of Library Networks and the Prospects for Their Development

The automation of library operations and the provision of services through a network does serve a valuable function in streamlining clerical work and allowing more effective use of staff, in improving services to patrons, and in rationalizing library planning and collection development for regions and states. The present disposition of library and information science both nationally and within the state indicates that some further advance in automation is desirable and feasible and that it should be done with an overall plan in view. The fragmentation of effort among libraries represents a waste of resources and in the area of automation creates difficulties for further development. The Federal government has expressed concern over this and will probably increase its support for state development of networks.

II. The Bibliographic Data Base and Library Applications

Although the ultimate direction of development nationally is not clear, states neither have to nor should wait before acting, because it is clear that the basis for any further development will be bibliographic records in machine readable form. States can begin to address this aspect immediately.

Since the central place in all library automation is occupied by the machine readable bibliographic record, the construction of a bibliographic data base should take precedence over any network effort directed to or limited by a single application. The beginnings of a bibliographic data base are already present in Virginia. The gathering and organizing of the already existing machine readable records could be immediately useful in the production of union catalogs and list. However, not all machine readable records are in a useable format. Programs will have to be written (or vendors contracted) to upgrade non-MARC records.

The conversion of manual records to machine readable form should be assisted and enlarged throughout the state and this effort should include retrospective conversion since it is essential to further development.

It is not necessary to begin a data base with records providing complete bibliographic information in MARC form. It is only necessary that records have enough information to be upgradable to MARC. Thus, a given system for which circulation was of primary concern could limit itself to contributing or creating brief bibliographic records for circulation purposes. Brief MARC records can be maintained with fuller MARC records in the same data base and even in brief form can be useful for applications requiring limited information (circulation, ILL, management statistics, limited information about acquisitions, etc.).

A data base constituted even of the present number and kind of machine readable records available in Virginia at the present rate of accretion, together with the importation of LC-MARC, would be sufficient to handle the critical mass of cataloging information needed by small and medium sized libraries and could serve as a base for retrospective conversion more economically than any available elsewhere. Larger libraries will require access to a larger data base than Virginia alone (even with LC-MARC) could ever hope to provide. These libraries should be allowed to join any of the larger interstate networks they choose, because this would not prove adverse to state development.

While replication of existing adequate systems is not advisable, for Virginia to build its own system need not be replication, because small and medium sized libraries might otherwise not be able to afford participation in a network and because certain applications will probably have to be developed at the local and state level (rather than by OCLC, BALLOTS, etc.).

Efficiency of search strategy and authority control have become major concerns in the automation of bibliographic information. Consequently, the development of adequate controls and search techniques should figure in Virginia's thinking from the beginning.

As far as particular applications, those needing attention first are circulation and inter-library loan. Systems should be designed with regional as well as statewide access in mind. Software can be easily developed to make the present and projected turn key systems compatible. Of systems available to the state, VPI & SU's Circulation and Finding System should receive special consideration because it is an adequate system able to be obtained more cheaply than vendor turn key systems.

III. Procedures Necessary in Further Development

The most serious obstacle to network development and planning is the lack of any organization or group able to carry such activities forward, even in a preliminary, exploratory way. At the present time, such an organization, effort, and dialogue are absolutely necessary before even tentative goals can be defined for further investigation. Before any specifications for the network can be drawn or estimates of cost given, the state must call together a working group for all sectors of the library community to help define procedures and immediate goals.

This groundwork is required because a statewide network can take so many directions that the only way of deciding the best option for investigation is by consulting with representatives (not just individuals) from the library community itself. Until this is done, "a State Library Network" will lack definition and direction.

There are presently 88 public libraries and public library systems, 87 academic libraries, 1,804 school libraries, and 143 special and institutional libraries in the state. The network is projected to address the needs of all. Obviously, this will not be done at once. It will have to be done very gradually, and direction and choice can be determined only when the library community has been sufficiently organized to provide the information about itself, its priorities and constraints, required for planning.

This articulation of the user community cannot be obtained by simply talking with individual librarians and proceeding on a one-to-one basis. It can be obtained only when individual librarians are selected expressly to represent their sector of the larger community and are specifically charged to work with others as a group designated to formulate requirements and goals valid for the whole group or defined in the context of a group. And we should stress again that the organization we are recommending is a group chosen to engage in preliminary activity. An organization for the network itself can be thought of only after the initial group has completed its work. The state must be prepared to undergo this process patiently, understanding that in projects of this size, years are normally required simply to formulate the fundamental considerations prior to any specific action.

Although the lack of this preliminary work creates major impediments in all areas essential to planning, certain of these should be specifically mentioned here:

1. No model can be projected either of specifications for network design or of costs because the specifications involved cannot be realistically determined. Any model projected would be nothing more than a shot in the dark because all parameters (i.e., the conditions which define the nature and results of any activity) remain merely hypothetical. If even mathematical modeling is to have any value, there must be some way of winnowing out the merely possible from the probable. This process must take place in the "real world" and not merely in the realm of mathematical hypotheses. In the case of a network, this expression of the "real world" must occur through a political process, i.e., by human beings discussing together their real situation and real choices for the sake of common action.

   The principal areas in which determinations need to be made are:

   a. Role the data base is to play

   b. The content and size of the data base

   c. Geographic distribution of users and potential users

   d. Density of traffic now and in the future

   e. Types of applications to be run against the data base now and in the future.

2. With regard to an analysis of costs:

   a. The state must designate the parameters for the options to be investigated, which can be done only after sufficient input from the library community

   b. When these are obtained, a general range of costs can be established for each option, but

   c. No adequate estimate can be made until the specifics of each project has been refined by a request for bids.

Throughout this report we indicate the parameters which must be determined for each area discussed. The State must now establish a working committee representative of the potential users to make the required determinations.

The State in implementing a network should move gradually, going from simple to complex. It should not commit itself to develop at the beginning a complete or "full-service" system, even as a pilot or model.

Certain areas in the state are more ready than others to engage in automating their operations and planning should allow for this. None the less, major attention should be paid to helping forward the less advanced systems, because the value of the bibliographic data base will be sharply curtailed if it is limited to only the more advanced systems. Although the State should encourage regional developments, it should at the same time engage in other activities that immediately address concerns which involve the entire state.

If the State Library is to play a significant role in fostering network development, it must increase its expertise in the technical considerations required. It should be prepared to acquire the expertise necessary whether by hiring appropriate staff or by availing itself of outside help.

A MODULAR APPROACH TO NETWORKING

As has been pointed out in the Findings and Conclusions section of this study, the library community within the Commonwealth as a whole is interested in some kind of computerized networking system. Opinions are varied as to what priority of use the system should provide. Areas such as Fairfax County, Tidewater, Henrico, Richmond and nearly all larger Public libraries and institutions such as VPI, Old Dominion and UVA feel that circulation, and along with it, inventory control and inter-library loan, are the most pressing problems. Of the 22 academic libraries now using the facilities of SOLINET, dissatisfaction with the quality of the data base, the searching capabilities of the system and the lack of any type of authority control is of prime concern. Those in charge of distribution of titles to the various public school libraries envision help in acquisition, cataloging and distribution of materials. Special libraries such as Phillip Morris desire a central on-line base of holdings in order to aid in acquisition and inter-library loans. The small public library is concerned with what role it could play within a network and, of greatest importance, what use it could be to them.

To be successful, a network must:

1. Provide useful service, of some kind, to all of its members;
2. Be developed in an orderly manner and in a way which avoids committing it to projects larger than its resources, experience, and expertise can support;

3. Be governed and managed with the interests of all users in mind and not just of a special group;
4. Meet the requirements of the user members in data content, response time, output products and management tools;
5. Be cost-effective and eventually self-supporting;
6. Be easily upgraded and changed as needed, both for state of the art technology and changing concepts of the library functions.

To accomplish all of this is certainly feasible. But how can it be accomplished in an orderly manner and yet allow for some "walking" and even "running" ahead by certain regions within the network? One of the findings and conclusions of this study, if not the most important one, is the fact that the whole area of governance and management for such a network has not been addressed. Findings from the study concerning the creation of a computerized network make clear that until it is known what a network is to be used for, what its data content and size will be and who will be using it, and how it will be funded (all are decisions that must be made by those governing the network), it is quite impossible to lay out any kind of actual design of such an entity.

Complicating the matter of design is the fact that some systems in the state are more ready than others to implement automation. These libraries have done a great deal of the preliminary work within their own areas, and are now ready to go ahead with solutions and do not care to wait for the rest of the Commonwealth to catch up. So the basic problem is how can the Commonwealth begin a networking process, one that will satisfy those who have not learned to "crawl" yet and also those ready to at least learn to "walk."

The following section addresses this problem and sets out, in modular form, an orderly progress toward a State Library Network as drawn from all available input to this study.

MODULE I - ONE TO TWO YEARS

Basic Facts:

1. The State has begun a base of holdings by contracting for a union list of holdings of the 22 academic libraries now using Solinet. These listings will be updated periodically and distributed to all interested libraries within the State.
2. Five areas (Tidewater, Old Dominion, Fairfax County, Henrico County, VPI) are at a state of development toward some type of automated system and they are ready to purchase hardware and software and are ready to set the plan into motion. (A good example of this is the study done for Tidewater and Old Dominion which should be available now.)
3. The State has not organized or set into motion any kind of governing body to make decisions about a network. What will it be and who will it's users be? Who will manage the actual facility? Who will make decisions about data base content, application programs, supporting system software, hardware support, etc.?

Action that can be taken:

As divergent as these three facts are, the State could begin a networking system with the tools now available, satisfy most of the library population, and most importantly be in a position to discontinue at the end of Module I, if it becomes the opinion of those involved in governing and network members that a fully automated library network is not desired, and still have a reliable and useful base of holdings.

The first action that should be taken is to begin the establishment of a governing body for the network. It must be emphasized here that this should not interfere with the other phases of this first module, but should occur in conjunction with them. The establishment of this group could take a great deal of time since in order for it to function properly it must deal with legislative problems involved in establishing a statewide network. The State library is in a position to begin this process. Care should be taken that all prospective members of the projected network have some kind of real input into the evolving rules and regulations set down by the governing committee. This does not mean that the group need be made up of all members of the network but that, at least, their interests and concerns are heard and some kind of action taken.

Once established, it will be the primary concern of this body to deal with the many facets of networking. It will be through this body and representatives of the network to it, that guidelines and rules will be established as to data content, authority control, network members, rules and regulations, network funding, site locations, future applications, and the many entities that go into the establishment of a network.

The Commonwealth of Virginia need only read and listen to the problems incurred and still not completely solved by OCLC and WLN to realize this process is slow and tedious and if not addressed in a manner that satisfies the majority of prospective users, dissatisfaction and dissension will occur, reducing the overall effectiveness of the network.

It would seem most reasonable to begin by forming some sort of intermediary body that would oversee and govern the activities taking place currently and who would study and produce an effective guideline of governance.

To accomplish this first phase of governance, the group should be representative of the various types of potential users (public, school, academic, special) but not so large as to produce an absolute stalemate which can occur in large special interest groups. It would also seem most logical that the State Library sit as head of this initial group.

Along with the establishment of this group to develop the guidelines for eventual governance, certain decisions concerning the beginning evolvement of the network must be made--mainly, those concerning the growth and direction of Virginia's Union list of holdings and the placement of regional circulation networks.

GROWTH AND DIRECTION OF VIRGINIA'S UNION LIST OF HOLDINGS

The State library has already commissioned a periodic publication in COM form of a union list of holdings of the State library and the twenty-two academic libraries participating in SOLINET and an individual shelf list of holdings for each member library. These are holdings cataloged since each individual library became a member of SOLINET and do not include, at this time, any retrospective holdings. The purpose of the union catalog is two fold:

1. To make available to each member library all materials held collectively and
2. To make the listing available to other interested libraries for original cataloging and inter-library loan purposes.

If the sources for the catalog remain limited to the present group, one can fairly ask whether the usefulness of the catalog will be correspondingly limited. One wonders just what purpose this listing will serve for the small and medium sized public libraries and for the public school libraries, and whether such a list sufficiently addresses the regional character of inter-library loan (e.g., the need for access to library holdings in one's own geographic area). If the catalog could be drawn from a greater number of sources, then, with regard to these questions, its usefulness would be enhanced. The next determination required, therefore, is how the basic data can be added to it without going into the costly business of communication networking.

Machine readable records provide the foundation for library automation and for networks based on automation. Available to Virginia are three groups of machine readable records which give bibliographic information pertinent to the holdings of libraries in the state as mentioned previously:

1. The archival tapes of the machine readable cataloging for the Virginia members of SOLINET,

2. The machine readable records which are comprehensive of various collections and which are produced and maintained either by a commercial firm or by the institution itself or an agency affiliated with it, and
3. The LC-MARC tapes which, although they do not provide information about which items are held in Virginia, do provide the bibliographic information required to catalog materials being received in the state.

These records are of varying kinds: some are cataloging records in MARC format, some are cataloging records in non-MARC format, some are abbreviated records designed to serve the needs of circulation, some are abbreviated records designed to serve both for circulation needs and as a brief form of cataloging to assist patrons in locating materials in the collection. The machine readable records of some libraries come close to encompassing all the collection, the machine readable records of others represent only that part of a collection acquired after a certain date (i.e., generally, from the time of their participation in SOLINET).

The tapes distributed by the Library of Congress provide the most authoritative cataloging information and are, of course, in MARC format. Tapes are available in different series corresponding to the different kinds of materials cataloged: books (monographs), series, films, maps, and government documents. They can be obtained for both current and retrospective LC cataloging (from 1968). In the case of books, 1,050,000+ titles were cataloged from January, 1968 to March, 1978. For current cataloging, tapes are issued weekly, containing approximately 3,500 records per tape. In the case of serials, 80,508 were cataloged from April, 1973 to March, 1978. Current cataloging of serials is issued every four weeks and contain approximately 3,000 records per tape. LC also distributes authority tapes for name and subject headings. Current cataloging of monograph includes CIP which is estimated to cover 80-85% of trade titles published yearly in the U.S. Although there was some effort to add records for materials prior to 1968 (through the RECON project), the vast majority of LC-MARC covers materials published after that date.

The SOLINET tapes contain cataloging information from the State Library and 22 academic libraries presently on-line to OCLC. The records are in MARC format and include the library's holding code. Archival tapes can be obtained from OCLC at intervals specified by the member institutions (although not more often than once a week). At present, the total number of records from all tapes is approximately 197,000, of which 100,000 are estimated to be unique. The ordinary rate of accretion for the present number of Virginia libraries on the system is estimated at 150,000 records per year. This does not include the number which will be added in various retrospective conversion projects--there are three now being planned through SOLINET: the State Library, the University of Virginia, and VPI&SU. The character of records on the archival tapes varies according to the differing cataloging practices of the institutions involved.

In addition to the records created and maintained through SOLINET, there are other groups of machine readable records which give bibliographic information about the holdings of Virginia libraries. These include:

1. Prince William County Public Library: For its COM catalog (maintained by Brodart), 77,000 titles in MARC format, with an accretion rate of 3,000/year.
2. Fairfax Public Library: For its COM catalog (maintained by Autographics), 250,000 titles in partial MARC format, with an accretion rate of 12,400 titles/year.
3. Virginia Beach Public Library: Now converting records for a COM catalog (maintained by Brodart), 90,000-100,000 titles in MARC format projected by March, 1979, with an accretion rate of 8,500-9,000 titles/year.
4. Henrico County Public Library: For its book catalog, 109,000 titles in non-MARC format. Henrico is in the process of installing the Cincinnati Electronic System which will up-grade records to MARC format.
5. Chesapeake Public Library: For its COM catalog, 70,000 titles in non-MARC format, with an accretion rate of 12,000 titles a year.
6. Chesterfield County Public Library: For its COM catalog (maintained by Autographics), 160,130 titles in MARC format, with an accretion rate of 7,680 titles/year.
7. J. Sargeant Reynolds Community College: 29,245 titles in non-MARC format, with an accretion rate of 7,000/year.
8. Northern Virginia Community College: For its COM catalog (maintained by Brodart), 100,000 titles in MARC format, with an accretion rate of 15,000/year.
9. Old Dominion University: Now completing conversion for its COM catalog (maintained by Blackwell/NA), 250,000 titles in MARC format (by the end of October, 1978), with an accretion rate of 19,200 titles/year.

10. Virginia Polytechnic Institute and State University: For its circulation and finding system, 250,000 titles in non-MARC format, with an accretion rate of 60,000/year.

11. Arlington: For its projected circulation system, 100,000 titles in non-MARC format, with an accretion rate of 11,000-13,000 titles/year.

The Virginia Union List of Serials and the Virginia Union List of Biomedical Serials have also been compiled from machine readable records. Although neither is derived from MARC format, plans are being developed to re-do the VULS and it seems probable that the standard adopted will be MARC.

The technology to convert these various machine readable records to full bibliographic MARC format is now a fact. This has been proven at Old Dominion which has converted its entire holding by running a condensed version of the record against Blackwell North American's data base and obtaining a hit rate of 85%. The same type of operation is being performed by Northern Virginia Community college. VPI&SU is in the process of converting its entire holding to machine readable form to be incorporated into its Circulation and Finding system. Its plan is to eventually process these records against a data base of OCLC records obtained from a similar university. There is no technical reason to believe that this approach will not be successful.

From these facts it can be seen that the original data base can be expanded almost immediately to include:

1. Public Library holdings
2. Community College holdings
3. Technical School holdings
4. Institutions of higher education holdings

The data base would now consist of a fairly representative collection of holdings within the Commonwealth. As each contributor to the original base adds unique titles and receives bibliographic information back from the vendors, they can be merged into the overall base of holdings, or, if the record exists, their holding symbol attached to it.

It would be up to the governing group to decide how the cost of such contribution by local and regional areas should be recompensed and how much non-participating members should contribute for use of the evolving data base.

The network would then be in a position to undertake:

1. The retrospective conversion of holdings of those libraries on SOLINET and the conversion of holdings of other interested libraries to machine readable code to be added to the data base and
2. The writing of computer programs to manipulate the data base for easier use, statistical reporting and management analysis.

Conversion Techniques

Although Old Dominion was successful in using a bibliographic base such as that provided by Blackwell North American, this may not hold true with the large research libraries within the Commonwealth such as UVA, William & Mary and VCU. Retrospective conversion of these holdings will more than likely have to take place through a larger and/or more comprehensive bibliographic system such as OCLC or BALLOTS. It is clear from discussions with some individuals involved that their primary concern in the area of retrospective conversion is time and personnel. Most are using all available resources just to keep up with the cataloging of new holdings. Help in this area needs to be forthcoming. It may be found that the use of another bibliographic system, such as BALLOTS, to do this retrospective conversion would be more efficient and cost-effective than continuing to do it through OCLC.

The newer institutions and those whose collections do not present formidable problems of range and depth could take the same course as Old Dominion or Northern Virginia Community College. This same type of conversion could be used for other libraries within the state who wish to add their holdings to the common base. If this approach is taken then the state, as a representative of the whole, should set up guidelines for conversion of holdings to machine readable code.

The pure logistics of such a conversion are horrendous and careful study should be made as to the approach and the vendor(s) to be used for the actual conversion and extraction of bibliographic records. To assume that each individual library could spare the personnel or absorb all the costs for such a conversion would be unrealistic. The state could facilitate the conversion process by:

1. Training teams to carry out the actual typing of the information to be optically scanned to produce a machine readable tape which could be run against a data base to produce full MARS records. The use of such a team or teams would no doubt speed up the conversion process.
2. Contract with a vendor of bibliographic data bases for the processing of these tapes to extract the correct record. The proofing of these could be done by the team or by a person designated by the individual library in question. The cost of conversion depends on a number of variables which can differ widely from institution to institution. Some, but not all, vendors provide general guidelines for costing. But unrealistic estimates can be made only when the specifics of the particular library are known. One thing is clear, however, the more conversions there are to do, the cheaper it becomes.
3. Providing for the creation of computer programs to enable the use of the already existing Virginia data base as the base these tapes are processed against. Even in its initial inception, after all forms of currently available machine readable tapes are converted and merged into the base, the "hit" rate could be quite high and would certainly not only be less expensive, but more efficient.

DATA BASE MANIPULATION

Once a base of some sort has been established, of what would be its usefulness not only to the contributing members but also to those who wish simply to gain access to the information available?

The production of a Union List of holdings and a shelf listing of holdings by contributing libraries would have been a product from the inception. However, as the base grows, this listing could become quite unwieldly and possibly could cease serving one of the purposes for which it was intended--that of inter-library loan. The base could be split out into regional groupings so that any library could easily determine what is available within its own geographic area without having to wade through the entire file. The development of this type of subset of the data base could also be of some use in the area of acquisitions. If the region is well supplied with material covering a certain subject, then the purchase of duplicate or similar material might not be necessary.

The whole area of statistical reporting and management analysis now becomes possible. Information can be provided regarding such matters as:

1. Number of unique title holdings and their distribution within the State;
2. Audio visual materials available and where;
3. Growth rate of new titles--in various disciplines and locales.

Before leaving the subject of the bibliographic data base, we should mention that the concern of all members of the library community (in varying degrees) is the purity of the data base, and the need for controls to ensure the propriety of the bibliographic record. This concern is legitimate and should be reflected in the planning and activity of the networks from the beginning and certainly in the event that the data base is used for on-line applications. One of the primary functions of any committee or authority controlling the data base at this point should be the scanning of the output from the various vendor's input material to ascertain its validity and coherence to standard cataloging practices, to set down guidelines and practices that must be adhered to by network members when and if the data is put on-line for new cataloging and to determine what kind of authority controls must be built into the system before it is put on-line for general use.

LIBRARY AUTOMATION STUDY IN VIRGINIA

As this study is being published there is in existence, within the Commonwealth, only one on-line circulation system, that at VPI&SU. There are several acquisitions systems which range from completely on-line to partially on-line to batch only. There are no systems which have integrated any of the various applications indigenous to a library.

In our survey of automated systems in the State, we have been interested in what is available now or in the near future and whether it will be feasible to include these components in a prospective library network. Before reviewing the various systems, it should be pointed out that we have been impressed at the care and thought evident in the choice and design of these systems and found them, for the most part, adequate for handling the basic operations concerned.

VPI/SU:

System:	Circulation and Findings
Computer:	Hewlett Packard 3000
Management System:	Hewlett Packard
Communication Software:	Hewlett Packard
Programming Language:	COBOL
Terminals:	Standard ASCII

As observed in use and as presented in discussion with library and data processing personnel at VPI, the circulation and findings system developed there seems to be quite viable. It was agreed by all that some changes should be made to the system, mainly in the area of serials control, but on an overall day-in and day-out use the system was doing the job intended for any such system.

The data that is captured by the circulation module of the system lends itself to being used as input, for extraction of full MARC records, from a bibliographic data base.

The programming language used (COBOL) makes it possible to convert the programs to any other host computer having a COBOL compiler. Changes would need to be made to the CALL statements to the Hewlett Packard data management system for whatever management was residing on the new host computer.

The Hewlett Packard 3000 system is of good quality and supplies all the software and hardware components needed in a networking system. (See NETWORK components: Appendix C.) Its cost, however, compared to other computer systems with the same capabilities is quite high.

VPI&SU also has developed a rather good acquisitions system. The system allows for on-line retrieval and updating of purchase orders, invoices, and tracking of the progression from purchase order to being on the shelf. (i.e., back ordered, ordered, received, cataloged, etc.) The system performs all the normal functions of accounting associated with purchasing and produces some rather sophisticated management reports.

Unfortunately, the system is not integrated with the on-line circulation system. Work is being done at VPI to determine if communication can be achieved between the Hewlett Packard circulation host processor and the IBM processor acting as the host for the acquisitions system.

Coincidentally, Fairfax School Media Center has developed a partially on-line acquisition system, written in COBOL, using a Hewlett Packard 3000 computer system and the Hewlett Packard data management system.

The group in charge of the direction the network will take, should certainly look into these systems and determine if it would be possible, and more importantly cost-effective, to make them generally available to the prospective users of such application(s) and the difficulty of making the two modules interactive.

Henrico County:

Vendor:	Cincinnati Electronics
System:	Circulation*
Computer:	UNIVAC V-77
Management System:	TOTAL
Communication Software:	Cincinnati Electronics
Programming Language:	COBOL
Terminals:	Standard ASCII
Costs:	$100,000 to $1.5 million

Henrico County has entered into an agreement with Cincinnati Electronics to act as one of the pilot installations for their circulation control system. As of the publication date of this study some phases of the circulation system were not complete nor fully tested.

An analysis of the system would be difficult at this time since the only persons to contact are the vendor or the parties interested in purchasing the system. If it performs as specified, the system should be capable of doing all existing functions of the VPI&SU system. As with all the turn key vendors in the field, Cincinatti Electronics plans in the future to add the various other applications desired (acquisitions, interlibrary loan, etc.) at a later date.

---

*Other applications as provided by the vendor

Pertinent to those making the ultimate recommendations as to the direction of the network and which vendors to use for various applications, Cincinnati Electronics will sell and support (fully if not modified by the user) the software alone. Observation of the system and its performance should lead to some valid comparative analysis for future purchases of this type of system within the network.

Tidewater Area and Old Dominion University

An extensive study on the development of a regional library networking system by Synectics Corporation is due to be presented to the Tidewater region and Old Dominion University concurrent with the publication of this document.

This group and Fairfax County are the closest to being able to establish some type of regional cooperative computerized network.

The Tidewater/Old Dominion directors were kind enough to let some members of this study group sit in on a preview of the forthcoming presentation by Synectics. It was obvious, from our stand point, that many of the difficulties that stand in the way of concerting a network development effort at the state level had been addressed and solved in that region. This does not mean in any way that the state has not thought about nor tried to address these same problems, it is simply that in a single region such as Tidewater, a comparatively small group of directors with common interests can solve problems much more easily than can be done on a state level where there are many involved and many interests to be satisfied.

The network as envisioned would eventually connect all interested libraries within the region to two or three host computers to initially handle circulation and other applications in the future. In addition, a microfilmed list of holdings would be published periodically for each library connected to the network.

As the catalogs are updated, the tapes will be processed against the Old Dominion data base and unique records or holdings identification added. From the data base, a union list of holdings for the area will be produced. It was not clear which circulation system would be used by the area. It is known that Old Dominion has decided upon the Data-Phase turn key system; however funds, although allocated, have not been released.

Old Dominion University is in the process of completing all its retrospective cataloging prior to becoming involved with SOLINET. It has now closed its card catalog and is producing its catalog of holdings in fiche form.

The next step in their overall plan is an automated circulation system.

Old Dominion

Vendor:	Data-Phase
System:	Circulation*
Computer:	Data General Nova or Eclipse
Management System:	Proprietary Version of MMIS
Communication Software:	Provided by Data General
Programming Language:	MUMPS
Terminals:	Standard ASCII
Costs:	System with 16 terminals, 192 million bytes of storage, 64K of main memory, magnetic tape unit and line printer: $200 - $250,000

The Data-Phase turn key system has taken great strides forward in sales in the past year. It is the first such system to pose any real threat (it is too early to tell about the Cincinnati Electonrics System) to the older more established CLSI system.

A search of the literature concerning the performance of the system and the general services offered by Data-Phase yields somewhat contradictory results. However, in the case of turn key systems this is not unusual. Often in promoting a new item, a vendor will promise applications which take much longer to implement than the user desires (thus, for example, although several of the major vendors have declared that acquisitions will be accommodated in their systems, only CLSI has actually achieved this and even here only for books--their serials check-in is still in the process of development).

---

*Other applications as provided by the vendor

As with Cincinnati Electronics and VPI&SU the system seems to handle well the normal functions one would expect from an automated circulation system, and the hardware and communications software is standard enough not to pose any serious problems if interfaced into a statewide network.

Fairfax County is another region within the state that has done a great deal of work and diligent planning toward automation of library services. Their study on library automation has been cited and used by many interested parties.

They were one of the first groups to close their card catalogs and convert all their holdings to machine readable code and produce microfilm versions of the catalogs and union lists.

Some while ago they submitted requests for bids on a circulation system network for the county. These bids have been opended (three were received) however, no decision has been made as of this writing.

From this brief analysis of the systems installed or projected as eventual components of any statewide network certain conclusions can be drawn:

1. All systems are capable of handling the automation of circulation
2. All systems utilize hardware components that could be easily interfaced into a larger network
3. All of the computers involved offer multi-processing languages (i.e., COBOL, FORTRAN, etc.). Libraries using these systems need not depend entirely on turn key vendors to add desired applications. The governing body of the network can elect to have these programs written to their own specifications. The installation and integration at each site might be different, but certainly not beyond the realm of feasibility.
4. All systems provide the basic information needed, in machine readable code, to allow the use of the circulation data base records for input in the retrieval of full MARC records which can be merged into the Virginia data base.
5. A relatively simple interfacing program could be written to allow these systems to communicate with each other and, in the case of an eventual on-line data base, with the combined bases to be used for inter-library loan requests and routing.

After the activities of the first phase in our modular approach have been implemented, those involved should have enough background and working knowledge to make a decision about the next step: to go ahead in the development of a completely integrated network or to stay at this level.

The legislation and governance of the network will have been established. Actual guidelines for the purpose of the network will have been set forth. All interested potential users will have been able to actually see and use output from the network.

By program manipulation of the common data base good statistics can be complied for the basic information needed as input before any network of this size and complexity can be designed. Answers to such questions as:

1. Number of transactions to be handled
2. Density of use in various areas of the network
3. Expected user response time
4. Back-up capacity needed
5. Size and complexity of the data base
6. Number of terminals needed and location
7. Parts of the entire system which can be distributed
8. Applications desired
9. Controls to be imposed on the common data base

If the state can supply valid answers to these points before issuing a request for a network design then it will be much easier to ascertain if the company(s) answering the request can supply the needed expertise to aid in the overall design and implementation of a network.

If the decision at this time is not to go forward with an integrated library system - nothing has really been lost and a great deal has been gained.

The common data base of holdings can still be updated periodically and union lists, regional lists, etc. can still be published.

A semi-networking system will be in place (various circulation networks) which can be utilized by other members in the community if desired. The ability these systems offer in data base search makes it easier to quickly answer any questions on inter-library loans. For comparatively little cost and effort, the systems could be inter-connected to further automate the inter-library loan function.

MODULE II - ON-LINE CATALOGING AND INFORMATION RETRIEVAL

After reading Appendix A "Library Applications" it becomes evident that an orderly progression toward a fully integrated library networking system assumes that a good data base of bibliographic records has been achieved and that guidelines and controls have been established amongst the members of the network to maintain the required quality.

Module I has provided the network with:

1. A common data base of holdings within the state;
2. The capability to update these holdings by various means;
3. Several output documents and materials from the base of holdings;
4. A governing body for the network.

The next logical step is to make the base of holdings available to members in an on-line situation for searching purposes and for cataloging.

It has been emphasized in this study that the need for adherence to standard cataloging practices and the use of good authority control of the file(s) is essential if the state wishes to maintain a data base of high quality full MARC bibliographic records. Those in charge of the decision-making about the content of the data base should make certain that the practices and guidelines are set down and agreed to by all current users before putting the base on-line. Potential users should be expected to abide by these practices also and these standards should be included in any contractual agreement.

By the time the data base is to be put on-line, much of the clean-up of the base will have been done.

1. Duplicate entries from OCLC
2. Decisions as to which cataloging to accept in case of duplications
3. Adding holding symbols

However, the base will consist of records from various sources and records which have not, in most instances, been subjected to any type of authority control. It is, therefore, essential that certain functions be completed before allowing use of the data base on-line:

1. The types of authority files needed should be defined and implemented.
2. A choice should be made of the "data base management" system that will be needed to effectively handle the current data base and also be useful in the future. This is of utmost importance since one of the most powerful tools the data base of holdings can provide is the quick and accurate retrieval of records. If the "data base management" system does not allow for dictionary or directory files, then one is pretty much locked into the use of key search such as utilized by OCLC.
3. A query module allowing for full Boolean arithmetic must be chosen.
4. File maintenance programs must be written to allow records to be added, deleted, changed and displayed. These programs should allow for as much editing and validation as possible. The MARC format is such that the computer program itself can, in many cases, force the user to input information that is mandatory and edit the input given.

5. Utility programs must be written to provide the kind of output materials expected by the users. During Module I the ability to produce catalogs in microform was accomplished, but there will be instances in which certain users still wish to maintain card catalogs. Provisions should be made to produce this type of output. Also, there may be users who wish to have their catalogs on-line to their patrons, again provisions should be made to allow this.

   Another utility program would be the so called "profile." This program, in essence, tells the system what functions this particular user wishes to use within the system and what security controls are to be placed upon various users.

When this is done and before any user is allowed access to the base, the files should be cleaned up. This may take some time and effort, but if not done before allowing on-line use, the whole purpose of the data base will be lost.

While the data base clean-up is being accomplished, the actual placement of the hardware and software components, as specified by the network design, should be going on. It is essential that good training in the use of the equipment and application programs be provided and that good manuals for users be provided.

When all preliminary preparation is accomplished and the network is on-line the base will contain:

1. Virginia common holdings data base
2. LC MARC data base
3. Authority files

The entire base must be structured so that one query will search all files in the order necessary as determined by the command given. The returned information must also tell the user from which file it was obtained.

For example, when cataloging a newly acquired volume, the Virginia base should be searched first and, if not found, then the LC MARC base should be searched.

The usefulness of such an on-line data base should be apparent:

1. Although certain academic institutions, special libraries, large public libraries will probably always need an on-line bibliographic service such as OCLC or BALLOTS, the majority of the libraries, especially the medium and small public libraries, public school libraries, community college libraries, will realize a very high "hit" rate by using the combination of the common data base and the LC MARC base. (NOTE: The updating of the file through the use of the archival tapes from the bibliographic services should be a weekly process.)

The costs that had been occurring, during Module I, of obtaining machine readable records for the data base from many libraries will now disappear. Cataloging can be done on-line in a real-time mode.

2. The ease of access to determine what is available will greatly speed up the process of reference searching and inter-library loans.
3. The data base will be current, with the exception of the cataloging done through a bibliographic service, but even these will only be a week old at most.
4. Users may now have any type of catalog desired:

- Card
- Com
- On-line

5. And lastly, it is now possible to begin adding other applications to the system.

MODULE III - LIBRARY APPLICATIONS

Again, we must allude to the governing body of the network, for it must be their task to determine what applications are needed, and in which order, to satisfy the needs of the majority of users of the network.

It could be assumed that if Module III is being considered at all, the network to this point has been a success and is deemed not only useful but cost-effective by the various users and by the state and governing body. In considering further applications, it should be remembered that to this point, use of the data base has involved applications more or less common to all members of the network. The basis for figuring expenses was accordingly common to all. But with an increase in the variety and choice of applications, the costs to libraries exercising this choice will increase. Therefore, each one of the application modules must be generalized enough in their design and conception to allow a certain amount of uniqueness to be given for the individual user without the cost being too high.

A hard look also should be given at this time to any applications that have been added to the turn-key systems already in the network. The reinvention of the wheel is certainly not cost-effective either.

Before beginning Module III, those in charge of guiding the network, along with current and potential users, should thoroughly study:

1. Types of applications desired by the majority of users
2. The order in which they should be addressed

At this point the governing group should engage an application design firm to study the entire network situation, to determine the design of the applications, to determine how they will be interfaced into an integrated total design, to look at and determine whether any of the application modules desired are already available on the commercial market, and if so, whether they can be purchased and used in the system, to determine the impact of the additions of the modules to the overall network (must there be an upgrading of hardware to allow for efficient use of the modules, etc.), to determine how and what information from the common data base may be used to formulate the records needed for the added application modules, to determine what changes must be made within the data base structure itself to accommodate the additional module, to determine where these bases should reside (on the central processor or on satellite processors), and produce an overall logical design, giving approximate costs and time frames to implement the desired modules.

To be able to accomplish such a task, the people involved (design consultants, governing group, current and potential users, data processing personnel) must be most cooperative.

Appendix A deals in some detail with the various applications relevant to the internal and external functions of a library. Therefore, in this section, we will only give an example of how one module (acquisitions) if properly designed and implemented, could reduce a great deal of the labor involved in the total library processes as well as acquisitions.

1. The design and implementation of the acquisitions module must allow for some front-end parameters (profile). If the network is to be used by all types of libraries, then the module must allow for:
   a. Sharing of acquisitions material and costs by certain pre-determined groups
   b. Total security of data, when desired, from other users as to acquisitions and costs
   c. The option of sharing information about items acquired, excluding costs involved
   d. Tracking system of where the volume(s) is in the acquisition, process by on-line change or by internal functions
   e. Easily changed (preferably by the user) reporting formats
   f. Integration with the on-line search and retrieval system for determining the following information by one access:
      (1) Does library already own the book and if so how many copies (catalog file or circulation file)
      (2) If not owned - has it already been ordered - if so - from whom; how many copies; where is it in the acquisition process (on-order file, in-process file)
      (3) If not owned - where else is it located? (common base of holdings)

g. When purchase order is cut for a given title, the system should automatically:

   (1) Update vendor file

   (2) Update invoice file

   (3) Update accounting controls

   (4) Set up circulation record (if library has automated circulation)

   (5) Set up cataloging record (its extent of information being controlled by the user)

   (6) Any other file that might be affected by the newly acquired volume should be updated from this one transaction. This could include the base of common holdings also.

h. When material is received the system should be able to:

   (1) Print any labels or materials needed including catalog cards

   (2) Update the common data base of holdings

   (3) Update accounting controls

   (4) Update purchase order and invoice files

   (5) And if desired by the user, issue payment.

There are many other functions that a good acquisitions module could accomplish. However, from just those listed it is quite evident that given an extensive data base(s) along with a good data base handling tool, many functions which would consume many hours if even done in a batch computerized mode, could take very little time in a fully integrated networking system.

This same type of saving in both costs and use of personnel can be found in the various other applications such as:

1. Circulation
2. Reference searching
3. Inter-library loans
4. Search control
5. Audio visual purchasing and materials booking

At the conclusion of Module III, the Commonwealth will have a fully automated and integrated library network answering the needs of most of the various types of libraries in the state. If the steps have been taken logically and each module has been designed and implemented with enough flexibility for changes due to functional needs and overall communication network load, the Commonwealth will be in a position, if desired, to offer the services to other states and/or groups.

The time frame and cost of such a network is impossible to determine until the steps included in Module I have been addressed and solved. The state of Washington has been working on such an effort for nearly eight years, spent nearly eight million dollars, and still has little but the on-line cataloging and data base retrieval in actual use. The University of Chicago has spent nearly ten years and as many millions. The facilities at OCLC and at Stanford required enormous effort and expense before benefits could begin to be realized.

The Commonwealth can avoid many pitfalls and circumvent foreseeable problems by making certain that whoever is involved in the design and implementation of each module has the knowledge and expertise to learn from the errors made by others who have attempted such a mammoth project.

A state network is feasible. Virginia could do it and, by doing it correctly and in an orderly fashion, make a significant contribution not only to the state and the libraries within it but to the entire nation.

APPENDIX A: APPLICATIONS

What should lie behind any concern with library automation or the automation of cooperative activities are the particular, practical needs to be met. What, then, are the specific library activities that can be directly aided by an automated library network? The account which follows will survey each of the principal areas of library operations, noting how that particular activity would be facilitated by automation, how it would serve in or be served by an integrated system (i.e., one which establishes a unified working relationship among separate functions--cataloging, acquisition, circulation, etc.), and what its relevance is for inter-library cooperation.

That area of a library's activity known as "acquisitions" can be regarded from two points of view: as a specifically bibliographic process and as a business/accounting procedure concerned with the purchase and receipt of library materials. In the latter case, its typical functions are: to create orders for items selected for purchase, to encumber appropriate funds, to issue claims and cancellations, to initiate or carry out procedures for payment (even if this means simply passing the necessary information to the office having authority for disbursement), to maintain information regarding vendors, to provide management information concerning types of material ordered and associated expenditures, to inform those requesting material of the results of their requests, to provide information on the status of materials in process (e.g., ordered but not yet received, received but not yet cataloged).

As it carries out these functions of ordering, receiving and accounting, acquisitions necessarily involves itself in the bibliographic process (i.e., that set of activities directed to locating records which describe items of print and non-print media). Acquisitions participates in this process for several reasons:

1. It must obtain information about available published material in order to select items to be acquired. (For the purpose of this discussion, selection, as well as pre-order searching, is considered integral to the acquisitions function even if it is not carried out in the "acquisitions department.")
2. It must check whether the library already has a given item or whether it is on-order or in-process
3. It must make sure that the information needed to order is correct and complete (i.e., correct author, title, publisher, date, identifying book number, price, etc.) as well as other information which concerns the item and which will be used by other departments of the library in processing newly acquired material (e.g., LC card number for cataloging)
4. When items are received, it must locate and update appropriate records.

In ascertaining the required information, acquisitions must consult a great number of sources and files: the library's public catalog, the shelf list, the desiderata, on-order, and in-process files, occasionally the circulation files (for missing or lost items), the serials file, the standing order files (for monographic series and blanket orders), vendor information files, and files generated by internal accounting. In addition, it must consult bibliographic tools for current (in-print) and out-of-print materials. Since there is no bibliography which covers all cases, the number of sources which need to be checked can be considerable. Simply for U.S. imprints, a typical (but certainly not exhaustive) list might include: the National Union Catalog in its various editions and cumulations, the Weekly Record, Books in Print, Publishers Trade List Annual, American Book Publishers Record, Forthcoming Books, Cumulative Book Index, Booklist, Choice, Ulrich's, New Serial Titles, and Union List of Serials. As one moves through other media and nationalities, the list of needed bibliographic tools grows to a length astonishing to anyone not familiar with the complexity of library operations.

But the bibliographic problems of acquisitions doesn't end with the number of files to be searched. Rather, in a certain sense, they begin there because the files have to be searched in a variety of ways and in various places. The variety of ways is dependent partly on the arrangement of the files and partly on the difficulties of determining the correct form and choice of entry for a given item to be searched. The problems of form and choice of entry which so notably confront the cataloger also, of course, confront anyone engaged in bibliographic retrieval. Consider, for example, the following hypothetical case:

> U.S. Department of Commerce, National Bureau of Standards. Proceedings of the 3rd Conference on Nuclear Reactors: Developments in forms of fission. edited by H. Swendenburg.

How should this be treated for pre-order searching? Is it a serial or a monograph (there might be different files for each)? How would one look for the author? Under U.S. or United States or Department of Commerce, or National Bureau of Standards or Swendenburg (presuming it is not Swendenborg)? And the title. Should one look under Proceedings? Conference? Under the specific title "Developments...? In the public catalog, one can look under both author and title. In the on-order and in-process files, under title only. In the NUC under main entry only. As this example should make clear, points of access, as well as number of files, are major problems in acquisitions.

The relation of acquisitions to other areas of the library's operations also needs to be examined. With regard to cataloging, not only does acquisitions use the public and shelf list catologs for its activities, but acquisitions is often regarded as a source of information useful to the cataloging effort. Since both are engaged in a search for authoritative bibliographic information, acquisitions can save the cataloger from needlessly repeating certain kinds of bibliographic searching by supplying him with the pre-cataloging information encountered in the process of pre-order searching. It can note, for example, the LC card number so that cards and copy can be ordered as quickly as possible.

With regard to public services, the acquisitions files are obviously useful in determining whether a requested item, not found in the public catalog, is on order, has been received, and, if received, its in-process status. The quick retrieval of this information and consequent speed-up in the processing of requested recent acquisitions can mean a significant improvement of service to patrons, especially for that group interested in recently published material.

Given these facts about some of the essential operations and relations of acquisitions, it should be possible to see more clearly the uses of automation in this area. Early efforts to bring acquisitions under automated control concentrated on the business operations aspect. Thus, systems were developed to automate purchase orders and claims, processing status reports, fund accounting and other activities consequent to the ordering and receipt of material. More recently, systems have shown that both the business and the bibliographic aspects can be successfully addressed by automation. In the BALLOTS system at Stanford, for example, it is possible to search at a single sitting, all of the required library files (catalog data, on-order, in-process) using multiple points of access in Boolean combination of key words drawn from both author and title fields. Further, the efficient relation of acquisitions to cataloging is considerably strengthened. The importing of LC MARC into the BALLOTS data base means, among other things, that items for acquisition can be searched automatically and continually against the weekly updated MARC file and complete cataloging data captured as a result. In non-conflicting cases, catalog cards can be produced at the time the material is received rather than having to wait until the material passes through the cataloging section proper.

One extremely important benefit of automation, not only for acquisitions but for any area of library operations has yet to be mentioned: its effect on file maintenance. In a manual system, the updating of files is a very time consuming and inefficient process. Ordinary human inadvertance together with occasional complications in the relation of sub-to-main headings or elements results in a rate of error that requires constant doublechecking and constant correction beyond that. For libraries with limited staff and ever-growing files, the problem is a source of continual frustration and occasionally the source of some very expensive mistakes. In a computerized system, file maintenance is automatic and takes place with invariable consistency. This is not to deny that machine filing has its own difficulties in accomodating customary human search strategies. Since, however, consistency is the chief concern in any filing arrangement, differences in the principles of arrangement need not be disastrous as long as the principles are understood by the user.

The implications which the acquisitions function and its automation might have for networking and library cooperation vary according to the way a given library sees its particular needs and the relation of its collection to that of others. If it is primarily concerned with the business operations aspect of acquisitions, then a network would be used to help such libraries either implement their own systems or participate in a central facility. But, as has been mentioned, that aspect we have been calling "business operations" is inevitably tied to the bibliographic process, and the history of developments in library automation indicates that control over the bibliographic process increasingly occupies the central place in the design of acquisition systems, both because of the natural importance of this process to the function itself and because of the integration it allows with other library operations. In such systems, inter-library cooperation is required because the retrieval of needed information depends on access to files of bibliographic information beyond those maintained by the individual library. At Stanford, for example, acquisitions uses the bibliographic data supplied by the Library of Congress and by members of the BALLOTS network. With regard to library cooperation, it should be noted also that in a network environment, the ability to share information about acquisitions among members of the system provides a basis for rationalizing the collection development of a state or of regions within a state.

The next activity in the technical services flow is cataloging. It is also the activity in which inter-library cooperation using automation and telecommunications has been most conspicuous. That this should be so is very interesting because cataloging, at first glance, would appear least amenable to automation. It is concerned with intellectual judgements of a high order regarding the form and content of a highly various group of materials. In making its determinations, it adverts to a very complex code of laws or rules governing every area of bibliographic description, intellectual and physical. What has been automated, therefore, is obviously not the cataloging activity itself but the exchange of information produced by the activity. This has been helped by the standardization of cataloging practice and the degree to which this had been achieved even before the advent of automation. From the 19th century on, librarians had recognized standardization of cataloging as a practical necessity. The bibliographic records of a particular library need to be accessible to an audience larger than that defined by the limits of a particular institution. Also, precisely because cataloging is demanding and time consuming, it became necessary to use centralized services which could supply catalog data to numbers of libraries which otherwise would not be able to keep up with the amount of material coming into the collections. The use of card services became common, and particularly significant was the establishment of the Library of Congress' card service division. LC became the authority and a focal point for national agreements regarding the conventions governing the form and content of bibliographic description. Apart from LC, the only other systems in general use have been the Dewey Decimal System for classification and the Sears list for subject headings. In the 1960's, the shared cataloging represented by the use of existing card distribution services began to suffer severe strain as a result of an explosion in the amount of recorded information. This meant that no cataloging source, including LC, could sufficiently keep up with the amount of material requiring its attention. It was also clear, and notably at LC, that the mechanism for retrieving, producing, and mailing requested cards was breaking down under its own weight. The huge size of the files and the immense demands on the service resulted in increasing inefficiency and increasing delays in responding to requests. In libraries across the country (particularly academic libraries), backlogs of books waiting to be cataloged began to reach alarming proportions. In response to this situation, new systems of shared cataloging have developed. The Library of Congress itself prepared the way by developing the standard format for the computerized recording and exchange of bibliographic information, the format known as MARC. The tapes on which the Library of Congress records its machine readable cataloging are distributed to interested libraries and organizations. The LC MARC tapes have served as a nucleus around which have been established very large data bases of shared cataloging (the pioneer and still by far the largest of these is OCLC). Appendix B provides a description of OCLC, BALLOTS, and the Washington Library Network.

Before moving to consider the specific uses of these large data bases, we should review the typical activities and responsibilities of a catalog department in a manual system. These can be grouped around three fundamental concerns: creating or modifying the bibliographic description, recording the bibliographic description for purposes of display and access, and maintenance of the appropriate files--the public catalog, the shelf list, the authority files.

The first of these--creating or modifying bibliographic description--is carried out either through original cataloging or by obtaining copy from some other source (e.g., LC proof slips, LC cards, CIP, NUC, a commercial vendor) and modifying details of it as required in particular instances. Whether or not cataloging is done originally, the cataloger adverts to the same set of considerations: the accuracy and completeness of the body of the description, the correctness and consistency of the choice and form of entries or headings, the appropriateness of the classification or call number. In order to make determinations regarding choice and form of entry, the cataloger relies on a set of authorities which normally include the national cataloging code (the most recent being AACR II), the list of approved subject headings generated by some commonly accepted authority (e.g., the LC list), and the library's own authority file of names and subject heading which are to be used for that particular collection. Decisions regarding the classification or call number rely on the classification schedules prescriptive for the general system, LC or Dewey, and on the particular disposition of materials in the individual library.

It should be evident from this account that questions of authority and consistency are extremely important in cataloging. It should also be clear that libraries subscribe to general codes but introduce a great variety of local exceptions and modifications. Consistency and authority mean walking a narrow line drawn at the intersection of local and common practice.

Problems of choice and form of entry deserve special notice. The code elaborated to assist the cataloger has grown more and more complex and in the process has itself become a source of misinterpretation and variety of interpretation. In a manual system, a cataloger unsure of the correct form might try finding authoritative copy in the NUC. But a successful search requires, of course, hitting on the exact form used. Since this is just what is not known, the cataloger might have to try one form after another pursuing it through numerous volumes, only perhaps to conclude at the end of the search that a particular corporate body (for example) was never entered there. Complicating the matter of authorized headings--and this applies to the classification schedules as well--are the revisions which take place constantly. Not only are entire cataloging codes revised but particular applications and individual headings are being continually changed and updated. Keeping track of these is extremely difficult; putting them into practice can be a major problem because the superseded entries in the catalog must be changed and cross-referenced. The trouble this can cause is well exemplified at the Library of Congress itself in its varying policies of superimposition and de-superimposition.

When the cataloger has determined the bibliographic description, records of it need to be made which can be used by the public and by the library's staff. This has normally meant catalog cards which a library either produces itself or obtains from some outside source. In producing its own cards, the procedure followed is to type a master or main entry card and then make copies of it for the shelf list and added entries. The additonal headings are then typed for that set. In the case of cards obtained elsewhere (from LC or a commercial source), if they do not come with the book, then a mail order procedure is followed. There can be considerable delays in response. Consequently, either a holding pattern is instituted for books awaiting cards or the books are briefly cataloged in order to get them on the public shelves. When cards arrive, the books are recalled from the shelves and fully cataloged. In any case, when cards arrive they are checked against the appropriate books, necessary modifications are made (which means typing or re-typing) and added entries are typed for cards in that set.

When the cards have been prepared and approved (and because of clerical error, double checking is necessary), they are ready to be filed in the public catalog and in the shelf list. Library filing is considerably more complicated than ordinary business filing because it deals with more than a single type of heading or access point, because it requires a much greater amount of subfiling, and because the order of significant elements is not always strictly sequential. The fact that the ALA rules for filing catalog cards constitutes a small book in themselves is some indication of the number of factors which must be taken into account. The procedure for filing in any particular library is determined in part by the way its public catalog is set up (i.e., whether or how it is divided). Filing normally takes place with batches of cards. This requires that they be sorted according to the type of catalog. If the public catalog is fully divided, this means grouping by author, title, and subject in addition to call number. They are then interfiled in each catalog. Because of the opportunity for clerical error, double checking is necessary before cards are lowered into place. Maintenance of the files does not, of course, end there. When books are withdrawn from the collection, when bibliographic descriptions are revised, when the records for any reason have to be changed, all of the cards for the item in question have to be pulled and then refiled. The process of catalog maintenance goes on constantly.

Another activity, not part of cataloging itself, but closely associated with it and often assigned to the cataloging section, is the preparation of the books for circulation. This involves typing and affixing exterior labels, book pockets and circulation cards bearing the identifying information necessary to locate the book and record its circulation. Also, if cataloging assigns the book to a special area within the collection (e.g., reference or reserve), the appropriate designation is affixed to the book.

From this review of the basic activities of cataloging, what can be concluded regarding the usefulness of automation in this area? It has been noted that the three principal concerns of cataloging are bibliographic description, the production of records for the description and the maintenance of the record files. With regard to the first concern, we have already touched on one major way automation has helped the cataloging process: it has made possible a more efficient system for sharing the cataloging burden by creating large collections of bibliographic information able to be rapidly retrieved for use by individual catalogers. It has also allowed the contribution of individual catalogers to be instantly added to the common pool for use by others. But automation can do more than organize large data bases for the rapid exchange of information. It can provide ways of more successfully coping with the search process, with the difficult problems of form and choice of entry, and with the problems of authority control. In systems developed after OCLC--at Stanford, WLN, Chicago, and Toronto--these other uses are being increasingly realized. In these systems, the computer can accomodate a search strategy using key words in any order, obviating the need to have the search formula correspond to the fixed order of entries. Variety of access has been further increased by the ability to search by subject. Also, name and subject authority files have been added to aid the cataloger in knowing and using the approved form and to prevent the duplication of records in the data base. In all of the data base facilities, the master bibliographic record can be modified to allow for local variations and cards produced according to the profile or specifications of each participant.

In the post-OCLC systems, time spent in finding authoritative copy has been reduced by instituting the capability of automatically searching the data base for information which had not been received at the time of intital inquiry regarding a given item to be cataloged. As MARC cataloging is added, the system automatically checks it for data pertinent tc the standing requests. It thereby saves the time and effort required to have a human operator engaged in repetitive searching.

The second major cataloging activity--the production of the physical records for bibliographic information--has also proved to be an area in which automation can usefully operate. Sets of catalog cards can be requested on-line from the central facility. This saves the separate effort and time involved in mailing out requests. And because the catalog copy requested is already edited for the individual library's use, the card sets obtained come with overtyped headings (and in filing order). Book preparation materials (labels and pockets) can be produced on site by a printer attached to the local terminal.

Machine readable cataloging also allows the production of catalogs in forms other than card catalogs. The archival tapes of individual institutions can be used to generate book catalogs in hard copy or microform. Computer Output Microfilm is an especially inexpensive method of catalog production and allows the library to have frequent cumulations and multiple copies for distribution.

The consequences of automation for catalog maintenance derive from these alternate forms of computer output. Book catalogs produced by computer eliminate the filing and human error associated with card catalogs. For systems which have introduced authority control into their bibliographic data base, revisions of name and subject headings can take place automatically without the long labor required to manually update the files.

In an automated system, the relation of cataloging to other activities in the library can be assisted and strengthend. The MARC tapes supply bibliographic information about current publications. As the tapes are received, they can be searched by subject to produce selection lists in any designated area of interest. Thus, the selection and acquisition process can be aided. We have seen earlier another way cataloging and acquisitions have been brought closer together--the use of the bibliographic data base allow acquisitions to anticipate the need for cataloging copy and to assume some of the burden of routine cataloging in instances of no conflict between authoritative copy and the book in hand.

Cataloging has an inherently close relation to the public service sections of a library because the catalog supplies the principal tool to be used by the public (and by the reference librarian) in gaining access to materials in the library's collection. Records in machine readable form can be retrieved selectively (for SDI purposes) or comprehensively in a format which provides improved scanning of entries--one of the advantages of a book catalog over a card catalog. Machine readable cataloging also provides the basis for eventual use by the public of terminals for on-line searching of a library's holdings.

Cataloging and circulation share the common concern of all activities in the library: helping the patron find the material he needs. Cataloging supplies circulation with the bibliographic information needed for identifying materials, information which circulation adds to or modifies for its own record keeping purposes. The interaction of circulation and cataloging can be aided in an automated system. If both functions are automated, an extract of bibliographic information for every item cataloged can be made available for the circulation records. Circulation can thus provide inventory information regarding the whole collection.

The circulation function itself is the next area to be surveyed here. Traditionally, circulation has been regarded as primarily responsible for the activities consequent upon the check-out and return of library materials. This requires that it maintain records concerning the borrower (name, address, ID no., status), the book (call number/ID no/author and title/category of material), and the loan itself (date of transaction, date due). It also issues overdue notices, collects fines, holds returned materials for patrons requesting it, and notifies them when it is available. In addition, circulation is generally expected to maintain and provide other kinds of information concerning the availability of materials--e.g., whether items are at the bindery, on reserve, in a special area of the library, in storage, missing, or lost. Circulation also provides management information about the use of the collection in its various sections and subsections.

Early in the history of computer applications, circulation was identified as a prime candidate for automation. Relative to other operations in the library, circulation's activities are more subject to the laws of clerical routine, use less complex bibliographic information, involve a greater number of transactions, require information more immediately (very often, while the patron is standing at the desk), and have greater need for up-to-date information (which involves more updating of its own records). Nonetheless, like all other areas of the library, circulation must manipulate a large number of files.

Automation has proven useful in improving the speed and efficiency with which circulation departments respond to the demands placed on them. Thus, pre-coding of information in machine readable form about books and borrowers allows the information to be recorded automatically at time of transaction without the laborious writing out required in a manual system and at time of return without the laborious manual searching of the files. Once the necessary data is entered into the system, it can be processed and out put according to the needs of the circulation function. A composite picture of the capabilities of various systems includes:

1. Provision of information on the location of circulating items--either all items, or only those items on loan or elsewhere than on the shelf (e.g., at the bindery, on reserve, being recataloged, or in storage)
2. Identification of items on loan to a particular borrower or class of borrowers (such as off-campus users)

3. Recording of "holds" or "personal reserves" for items on loan but desired by another borrower, often with additional provision for notifying the library staff when the item is returned and printing a "book available" notice for mailing to the person who requested the item

4. Printing recall notices for items on long-term loan

5. Renewal of loans

6. Notification of the library staff of overdue items and printing of overdue notices

7. Notification to the library staff of delinquent borrowers (i.e., those with unpaid fines or overdue books) either at the time of an attempted loan, at the time a borrower is leaving the institution, or on request from the library

8. Calculation of fines, printing of fine notices, recording receipt of fines, and sometimes printing of fine receipts

9. Calculation and printing of statistics of various types

10. Analysis of both summary statistics and statistics for the circulation of particular items for use in acquisitions, planning of services, and for other administrative purposes

11. Provision for handling special categories of borrowers and special types of materials

A few systems also print date due slips, automatically generate orders for lost books or needed additional copies, and print mailing labels for remote borrowers.

Developments in automated circulation systems are clearly in the direction of on-line systems--not only because of advances in technology but because of the need in circulation to make adjustments and corrections on the spot and to have access to updated records as soon as possible. These on-line systems provide CRT display, terminal printers, and data entry by means of bar-codes or OCR labels scanned by light pens.

As automated circulation systems have developed, a change or shift of emphasis can be observed in the definition of the circulation function. This change is very significant for the interests of this study. Circulation is being viewed increasingly as an inventory/finding system rather than simply an "absence" system (i.e., it is considered as a source of location and use information about all the items in a collection, not just those out on loan). The potential for circulation to act in this expanded role was always present because of the nature of the function and its place in the order of library operations and because of the variety of tasks which from the beginning had been "added" to its main activity. But the constraints of manual operations and the lack of integrated systems meant that circulation had time only for the most pressing demand--keeping track of check-outs and returns.

The consequences of their new emphasis for inter-library loan will be discussed later under that heading. For in-house operations, the change will mean a much closer working relationship between circulation and the other departments of the library. In a fully integrated system, data which previously could be found only in the separate records of acquisitions, cataloging, and circulation will be merged to provide from a single query station complete information about the status of any item relative to the library's collection (whether it is on order, has been processed, or is immediately available).

The next area of public services to be examined here, reference, is of all the library operations the least involved with creating and maintaining records and files. But in its use of them, in the variety of records it must consult, reference surpasses all other areas. The range of its interests is greater since it must be concerned with the whole universe of knowledge, not just that part of it contained in or appropriate to the immediate library. Reference deals directly with the patron in helping him locate the information he needs. It must turn to (or point the patron to) the various guides, bibliographies, indices, catalogs, and files appropriate to a given field of inquiry. The variety of organization, format, and search strategy among bibliographic tools even for the same field is very great. The labor of manually searching them requires moving from index to index, and within each index from volume to volume (year to year), and within each volume from significant term to significant term. At each turn, the process is complicated by the same problem that affects the search process elsewhere--form and choice of entry.

The reference process is concerned with both bibliographic and location information and deals with two types of bibliographic tools: those which give bibliographic information without adverting to whether or where the items might be in the immediate library, and those which do provide information about what the immediate library holds and where it might be found (union catalogs can be regarded as falling in this second class since they provide information about the holdings of a group of libraries). For the latter class, we have noted the way in which an automated system can allow information about the library's acquisitions to be more efficiently accessed. To the extent that reference personnel are able to use these systems for their own needs, they can be of benefit in the reference area.

For the first group of bibliographic tools, automation has also been of use. Very large data bases (e.g., Lockheed's Dialog, SDC, BRS, Medline) have been constructed which bring together the major indices to current literature, books, journal articles and technical reports of various fields. These can be queried on-line and the resulting citations and, in some cases, abstracts printed out for the searcher. Their chief advantage over a manual system is that they allow many indices to be searched cumulatively and simultaneously using multiple terms in Boolean combination. This means a radical reduction in search time and a radical increase in the power and refinement of search strategy. Further, the information searched is more current than that available in a manual system. Indexing information can be entered automatically into the data base and made available for general use in much less time than that required to produce and distribute hard-copy cumulations of an index. Since the audience aimed at in these IR services is primarily a professional and technical one, working in fields where currency of information is very important, the timeliness of citations is very important.

Libraries that wish to gain access on their own premises to Information Retrieval facilities must acquire a compatible terminal, printer, and communications equipment. Charges for use of the services vary: in some instances fees are based on actual use only, in others, they are based on a standard monthly charge plus actual use. The cost of connect time depends on the particular data base being searched.

The use of such search services has several implications for inter-library cooperation in regard to costs, staff training and time, and support materials. Although the cost of equipment and connect time have been steadily decreasing at a significant rate (the cost of use has been reduced by 50% since 1975), the expense might well continue to be too great for individual libraries to bear, particularly for those whose kind of clientel is not likely to generate a demand for service sufficient to justify the investment. Further, in considering the ability of a library to support these services, factors other than the direct costs of equipment and use must be taken into account: overhead, staff time, the need for training, and the need to achieve and maintain expertise in order to keep the search economical and efficient--and efficient use of these facilities does demand special training, a staff who must interact with the patron, and an on-going familiarity with the various search techniques and terms required. Individual libraries might not have the staff, the time, the money, or the demand. For libraries who would like to offer this service to those of their patrons needing it, but who cannot afford to support such service on their own, inter-library cooperation can offer an acceptable solution, particularly in geographic areas of high population and a multiplicity of libraries and library systems. The designation of a central facility, supported in part by the participating libraries, to which patrons at the various libraries can be referred, might secure a sufficient volume of work to justify the special staff, training, and equipment required.

Thus far in this survey, we have examined cooperative activities involving the exchange of bibliographic information. The inter-library loan activity involves the exchange of the materials themselves to which the bibliographic information refers. Inter-library loan must go through a verification process similar to that required in acquisitions. It must check to make sure the library does not actually own the requested material and it must check the request information for accuracy and completeness. Then it must forward the request either to a central facility or to the appropriate library (which means checking various union lists and guides to other collections). The request is normally sent by mail, telephone, or teletype. The receiving library then checks its collections to see whether the requested item is available. If it is, the lending library must prepare the material for delivery and send it out. If it is not, then it must notify by phone, teletype, or mail the requesting library or agency. In this event, the requesting library can then continue the search until it finds a library able to fill the request.

The speed and efficiency of the process is affected by the following factors: the time required for bibliographic search (this is affected by the amount of information needed to fill a request, and the adequacy of union catalogs and location lists), the time required to notify a library having the material available (this is affected by the availability of circulation information in addition to location information, and by the adequacy of a routing system which identifies the likeliest resource centers for a given region and which can distribute, where desireable, the burden of lending activity), and the time required to respond (this is affected by the time required to search the shelves and prepare the material and the time required to deliver).

Automation and telecommunications can significantly affect the total time required and can simplify the processes involved. Bibliographic verification can be done by access to the same data bases used in acquisitions, cataloging and reference. Libraries holding the material can be located either by on-line inquiry or off-line thru the union lists produced for libraries on the system. The potential lender can be notified through terminal-to-terminal communication. If the circulation system is automated and integrated into the network, the borrowing library can be apprised on-line of the circulation status of the needed item at the library being queried. Further, a routing system can be designed both to distribute the lending burden and to continue searching until a successful reply is obtained. Also, such a network would be able to support the extension of inter-system borrowing privileges to patrons as well as to libraries.

APPENDIX B: BIBLIOGRAPHIC NETWORKS

In the following description of non-commercial library network data base facilities, we have chosen those which are in actual use for networking (rather than in a status of potential use) and which are established in the United States--this last qualification is added because of the special problems associated with a non-U.S. facility, with regard to both the design for an eventual U.S. national network and the expenditure of federal and state funds. The three systems which will be reviewed are the Bibliographic Automation of Large Library Operations Using a Time-Sharing System (BALLOTS), the Ohio College Library Center (OCLC) and the Washington Library Network (WLN). The BALLOTS system was designed and implemented at Stanford University,Palo Alto, California. It has been providing on-line technical processing services for the Stanford University libraries since 1972. OCLC has its central office at Columbus, Ohio. It was founded in 1967 by the Ohio College Association in an effort to increase availability of library resources and reduce the rate per unit costs in libraries. It has a regional office in Claremont, California, to serve its western region members. Seventeen regional networks representing more than 1,056 institutions are responsible for profiling, user training and maintenance of OCLC services in their areas. WLN has its central office in Olympia, Washington, at the Washington State Library. Member public libraries throughout the State participate in the network. In 1978, WLN, with the addition of libraries in Alaska, moved to an active interstate status.

The three networks vary in number of members. OCLC is the largest with more than 1500 participating libraries. BALLOTS has 102 participating libraries and WLN has 20 member libraries. Of BALLOTS' 102 members, 45% are academic libraries, 16% are public libraries, 20% are special libraries, 8% are Federal/State libraries and the remaining 11% are grouped as other libraries. OCLC has 1203 profiles. (A profile is not necessarily a single institution; a branch library may or may not be a separate profile or a single profile may include several separate libraries in a processing center arrangement.) OCLC's member libraries are 71% academic, 11% public, 6% special, 8% Federal/State, 4% other. WLN's member libraries are: 40% academic, 50% public and 10% are Federal/State libraries.

Search techniques vary from system to system, although there are some strategies common to all. Through BALLOTS, searching can be done using several techniques. A combination of index terms from several indices is used to retrieve the desired record. The most common form of search is the personal name title search. An example would be "find pn Logan, T and Ballots view." This search would retrieve all records containing the personal name "Logan, T" and the title words "BALLOTS" and "View" will be retrieved.

Other search requests that can be used are: corporate/conference name, the unique BALLOTS record identification number, subject, call number and LC card number. Boolean operators of "and," "or," and "and not" are used to make a single criterion of two values. The truncation symbol is also utilized to retrieve records where values begin with the characters before the "#" truncation symbol.

BALLOTS Catalog Data File allows each library to catalog their own distinct entry and indexes these as required by the form of cataloging. The Catalog Data File, comprised of sub-files, can be searched separately and in any sequence.

The OCLC System uses seven search keys to access the data base. The keys are author/title, author, title, LC Card Number, ISBN, ISSN, CODEN, OCLC Control Number.

OCLC recommends the usage of the LC Card Number, ISBN, ISSN, CODEN, or OCLC Control Number as being the quickest methods of retrieving a record. The derived search keys are not as desirable to use because they have a longer response time and often retrieve several records.

The author/title, author and title search are performed using an abbreviated form of the data. The author/title search is a 4,4 combination using the first four letters of the author's last name and the first four letters of the first non-article word in the title. Thus, to retrieve John Gardner's October Light "gard, oct." would be the search key.

The author search key is 4,3,1 combination, using the first four letters of the last name, the first three of the first name and the middle initial. The use of the author search is limited to 7 - 9:00 a.m. and 5 - 10:00 p.m. Eastern Standard time, because it places a heavy burden on the system and its use degrades response time. The title search key retrieves all title entries in a similar manner using 3, 2, 2, 1 construction.

Searches using corporate author and author/title can be limited using a "Stop List" which contains terms that often appear in corporate names. This allows the search to use more specific terms and retrieve fewer records.

Another way to limit the search is by truncating the search key. When a word contains fewer letters than allowed by the search key, the key can be limited to the length of the word by the use of a circumflex.

When more than a single record is retrieved, those records are listed in abbreviated form. The user selects the desired and the full record is displayed.

The OCLC system limits the number of records that can be retrieved to 256. If more than 256 are found the search must be conducted using a different search key.

The OCLC data base consists of a single file to which all libraries add their cataloging. There can be duplicate records and when duplicates are encountered in a search the user must look at both records and determine which one to use. At present, duplicate records are not deleted, but the OCLC staff is improving the system so that it will identify duplicates, delete them and switch library holding symbols to the remaining record.

The WLN Bibliographic files are searched through author, title, subject, ISBN, ISSN or LC card number. Titles and corporate or conference names may be searched for any number of key words entered in any order. Headings from the authority file may also be used to retrieve linked records from the Bibliographic file.

The authority file contains name, subject, and series headings that are under authority control and also related cross references. Headings from the authority file are retrieved using the same author, subject and series searches that are used with the bibliographic file. The authority file may be searched using specific types of authorized headings. For example, it is possible to search corporate name series headings, topical or geographic subject headings, or personal names used as subjects. These headings in the authority file are linked to the records in the bibliographic file which use them. Thus, a user can verify a heading using the authority file and then request a display of all bibliographic titles using that heading.

The holding file records indicate which libraries in WLN own a bibliographic item. The holding file may be searched using LC card number, ISBN and ISSN, but more frequently it would be searched by referring to the immediately preceding search of the bibliographic file.

The searching of the WLN system can be refined or expanded using Boolean Operators "and," "or," and "but not" with search arguments to broaden or narrow the scope of a search. Using Boolean operators it is possible to search for an item that deals with both urban renewal and architecture. Truncation is used to broaden the scope of a search. The first part of a search key is entered and then a special character is added to indicate that this keyword is a shortened or truncated value. When the truncated value is used, all items with access points matching the search argument will be retrieved. For instance, the truncated subject "economic" will yield economics, economic forecasting, economics in literature, etc.

The WLN system also allows the user to search using subdivisions of MARC format authorized headings. For example, MARC subdivisions of personal name include: name, numeration, titles and dates. Any one of these subdivisions can be used as a search value.

The hours during which the computer services are available are:

1. For BALLOTS 8:30 a.m. - 1:30 a.m. Eastern time, Monday - Friday, with a break from 9:00 - 10:30 p.m. Eastern time, M, W, F. On Saturday from 4:00 p.m. - 1:30 a.m. Eastern time.
2. For OCLC 7:00 A.M. - 10:00 p.m. Eastern time, Monday - Friday and 8:00 a.m. Eastern time on Saturday.
3. For WLN, their hotline is available from 11:00 a.m. - 8:00 p.m. Eastern time, Monday - Friday.

Downtime must be expected in any computer network system and each of these systems has experienced downtime. BALLOTS reports its downtime as being 5-6%. OCLC reports its downtime as about 9%. WLN has recently installed an Amdahl 470 V611 computer which they feel has greatly increased their system stability.

The response time for BALLOTS has averaged three seconds during a normal day and six seconds during maximum load.

Average response time for OCLC has been from 4.9 - 8.5 seconds during peak hours of operation. The response time may be longer due to number of terminals on the phone line, time of day and type of request being made.

The response time experienced by WLN has been typically under four seconds.

The amount of time needed to start production operation with BALLOTS varies as to whether you desire the line by line or full face mode. The line by line method requires one to two months, and the full face takes from two to three months.

With the OCLC system full set up time for a brand new participant using high speed communication takes four months, while a new member using a dial-up terminal or a library sharing an existing terminal can become operational in two months. Start-up information for WLN is not available.

Basic cost figures for the three systems are:

	BALLOTS	OCLC	WLN
Terminal Purchase	$3400-$3700	$3700	$3500
Installation	$100-$200	$148	$75
Terminal Maintenance	$40-$60/mo	$33/mo	$35/mo

BALLOTS online data base, which contains over one million records, offers users the capacity to search the files, input standing search requests, catalog on-line, produce cards and tapes, produce on-line data base entries and request replacement cards and tapes.

BALLOTS services are offered at various levels so that a user may elect to participate in just a few of the services and pay for only those services used. The products that are available are catalog cards, tapes of their cataloging records, input worksheets to check and revise the library's cataloging. Standing search requests are available so that an unsuccessful search can be automatically re-tried at a later date. BALLOTS bibliographic records contain location symbols that can be used as a source for inter-library loans. BALLOTS also offers consulting and user training as necessary.

The OCLC on-line data base, which contains nearly four million records, offers numerous bibliographic services to its users. These services include an on-line union cataloging accession list, book pocket and spine labels and automatic check-in of serials.

From the on-line cataloging, catalog cards from all records can be generated, magnetic tapes of a library's catalog holdings accession list, book pocket and spine labels and automatic check-in of serials. OCLC bibliographic records contain location symbols that can be used as a source for inter-library loans. Start-up instructions, training and follow-up support services are also offered.

WLN's on-line data base contains over one million records and offers on-line input and editing of bibliographic records. Products that can be generated from the bibliographic records include catalog cards and spine/pocket labels. Customized book catalogs may be produced for one library or group.

Future services to be offered by BALLOTS include a multi-library acquisitions service similar to the one now in use at the Stanford University Library, a multi-library cross-reference file service, increased searching capabilities, interfaces with other automated library processing systems, support for other MARC formats besides Book Authority control of records and inter-library loan.

OCLC's announced future services are to include: serial cards, inter-library loan, acquisitions, subject retrieval, serials control, authority control, circulation control and monographic local headings.

WLN's future service which are now under development include an acquisitions sub-system to monitor acquisitions. A circulations subsystem is under development by Data Phase (circulation records, check-in/out, overdues, fines, etc.). A serials control subsystem is also under development.

APPENDIX C: NETWORK COMPONENTS AND STANDARDIZATION

COMPUTER NETWORKS

One of the findings and conclusions of this study deals with the reasons a detailed network design cannot be presented at this time. In essence, the conclusion states that basic information needed as input to such a design is not available nor will it be available until certain functions leading to a design stage have been reached. (Module I - Logical Steps Toward Networking).

Instead, this section of the study will deal with the hardware and software components that should be available in any state of the art network and with what is being done regarding network and communication standardization. This, therefore, can act as a document to help people not schooled in data processing techniques to at least be initiated into this complex and comparatively new development in the sharing of computer information.

A major force in society's growing use of computers has been the increasing volume of transactions to be processed. For instance, in the mid 1950's, Bank of America pioneered the use of computers, due to the realization that in the future the manual handling of check would require the entire population of California. It is clear by now that if computers were not available, the level of activity in many segments of society would be limited by the fraction of the population that could be devoted to the varied bookkeeping tasks. It is only recently, however, that it has been realized that a similar limitation is beginning to emerge with respect to tasks of a substantially higher intellectual character.

Many complaints are heard (and in the case of California's much reported Proposition 13, action taken) regarding the inadequacies of medical care, of education, of libraries, and many other services. These inadequacies are blamed on a variety of factors such as incompetence, poor planning, poor use of available funds or resistance to change. It may be, however, that these factors are merely symptoms of a more fundamental limitation. Namely, that society does not possess the human resources necessary to perform adequately all the necessary tasks, particularly in view of its growing complexity and the rising expectation on the part of the population.

To provide some kind of solution to these problems, computer networks are fast becoming a powerful national force and are exerting influence on all segments of society.

While the use of sophisticated jargon by telecommunications specialists makes them seem modern and formidable, in reality networks are not new. The rail system is an example of a network used for commodity distribution. The electric utility companies operate similar networks for the distribution of electric power. The national telephone system may be viewed as a network for the distribution of information.

It is the information distribution type of network which serves as the model for computer data networks.

Data networks provide a means for sharing expensive computer resources such as data bases, specialized software (computer programs) and the hardware itself. Yet, until recently, it was necessary to make substantial investments to construct private networks thus restricting the benefits of networks to specialized communities (such as airlines or banking) or the large corporations who could afford such an investment. There were no public facilities in the sense that the railroad network or telephone system is public.

Recently the situation has changed. Pioneered both by government research and private initiative, a new concept in networking has emerged--Value-Added-Networks (VANs). These systems build on the already existing communications facilities to provide a new kind of public network service. A VAN utilizes the existing common carrier network for data transmission and provides the additional services of interfacing the users' computers and terminals, routing messages to their destination while guaranteeing their integrity, and allowing small users to take advantage of the economics of scale inherent in large communications systems by pooling their demands for service.

COMPONENTS OF A DATA COMMUNICATION NETWORK

Hardware

In this section we will address the hardware components from which computer networks are contructed.

An Overview of the Data Path From User to Host

The data path from the human computer user to the servicing host through a computer network is a complicated trail beginning with the local terminal. The terminal is connected to an adapter which modifies signals to a form acceptable to the local transmission line connecting the terminal to the local network port. If the terminal and network port are located closely enough, this adapter may not be necessary. Another set of transmission facilities interconnect network ports. At the serving host, another network port is required, along with a hardware connection between the host and the port. If the host and port are not contiguous, one or more adapters and local transmission lines may be required.

Circuits

The circuit is one of the basic building blocks used to construct a network. There are those circuits that interconnect the nodes of the network, circuits that connect terminals with the network, and those circuits which provide the end-to-end path between communicating processes. Consider first the circuits that physically connect the nodes in a network. Available circuit capacity ranges from 110 BPS for voice grade lines to 250,000 BPS for wide-band. The second kind of circuit in a network is that connecting a customer device - timesharing terminal, remote job entry station, minicomputer based system, etc. - with the network interface, possibly a multiplexer or concentrator. This circuit, part of the low speed channel side of the multiplexer or concentrator, can operate at a maximum speed equal to that of the circuit connecting nodes, but is usually sharing that high speed circuit with a number of other low speed circuits and thus operates at a considerably lower rate.

Competition and technical innovation are currently changing the variety of circuits available to the network builder. Regulation and litigation further complicate the offerings. Furthermore, specialized common carriers may include as part of their offerings such features as error control which presently must be implemented by the network.

Carriers

The established old line common carriers in the U.S. include the local telephone companies, A T & T long lines, and Western Union. Newer carriers offering, planning to offer, or awaiting approval to offer service include United States Transmission Systems, Inc., a subsidiary of ITT World Communications; CPI

Microwave, Inc.; Data Transmission Company (Datran); MCI Telecommunications Corp.; Nebraska Consolidated Communications Corp. (N-Triple-C); United Video; Southern Pacific Communications Corp.; Western Tele-Communications (WTCI); and American Satellite. The newer carriers do not offer uniform geographic coverage.

Interactive Terminals

Interactive terminals are generally equipped with a keyboard and an incremental printer. Optionally they also include a tape subsystem (paper or magnetic). A display may replace or complement the printer. Such terminals support interactive, conversational, demand, inquiry, and transaction oriented applications.

Information Transfer Rate

One important description of the data path, including terminal adapter, or "modem", and transmission line, is the maximum rate at which it can transfer information. Before identifying two such measures, it must first be clearly stated that the raw transmission capability is not the same as the net rate at which information is transferred. There is always an overhead. The maximum information transfer rate is expressed in bits per second (BPS). A bit is a binary digit, the smallest unit of information. The more meaningful, but less used, measure is the transfer rate of information bits (TRIB), which excludes all overhead. Another descriptor commonly used is baud, the number of state changes per second. If there are multiple states, then each state can represent more than one bit; therefore the BPS is greater than or equal to the baud. For example, a transmission line capable of supporting four states has a BPS rate twice its baud rate since each state represents two bits.

Analog/Digital Adapters

When circuits designed for voice communication are used for data transmission, there must be a transformation between the domains. The adapter between the terminal and the transmission line provides modulation and demodulation. The name "modem" is a contraction of this functional description.

The Function of a Modem

The modem performs a transformation between the digital signal acceptable to the terminal, network port, and the host computer and the analog signal employed for transmission. In the modem, the square-edged pulse train from the computer or terminal is transferred to fit in the telephone channel frequencies between 300 and 3300 Hz.

Modulation Techniques

Transmission rates on the order of 188 BPS are the present maximum available from FSK on voice-grade lines. Amplitude modulation (AM) varies the amplitude of the line carrier frequency. Frequency modulation has a noise advantage over amplitude. Phase modulation (PM) in the form of differential phase-shift keying (DPSK) is employed for higher speeds; 2400 to 9600 BPS have been achieved on conditioned lines. The term differential implies that symbol meaning is based on a change in phase from the previous state and not on absolute phase reference. Four states of phase shift are commonly used, producing a BPS of twice the baud. Increasing the number of phase shift choices to eight yields a BPS of three times the baud.

Asynchronous Modems

In asynchronous mode, one character is sent at a time, preceded by a "start" signal and terminated by a "stop" signal. FSK modulation is usually employed; therefore, the BPS and baud are equal. Since the transmission rate is determined by the digital signal, it may vary up to the maximum supported. The existing conventional rates are 110, 134, 150, and 300 BPS. Provision may exist in the character code (e.g. ASCII) for error detection in the form of character parity.

Synchronous Modems

In the synchronous mode characters are sent in a continuous stream, thus requiring three levels of synchronization: bit, character, and message. Transmission is synchronized by a clock internal to the modem. This fixed rate transmission does not require start and stop bits.

Protocols

Protocols are required for bit synchronization, so that the receiver knows when a bit starts and ends so that it can be sampled; for character synchronization, so that the receiver can determine which bits belong to a character; and message synchronization, so that the receiver can recognize the special character sequences which delineate messages. Typical protocols include the blocking of transmission into messages employing start-of-text and end-of-text (STX/ETX) or other similar markers and a positive/negative acknowledgement procedure (ACK/NAK). Additional error detection and correction is provided by longitudinal redundancy checks (LRC).

Conditioning/Equalization

The analog signal loses strength as it travels along the transmission circuit. The loss is proportional to frequency. Likewise phase shift occurs also as a function of frequency. Attenuation and phase shift are components of distortion which can be compensated for by adding complementary filters. When the filters are added to the physical plant, the process is called conditioning. When provided within the modem enclosure, it is called equalization. For leased lines fixed or manually adjustable, conditioning or equalization suffices. When switched circuits are employed, automatic equalization is a major convenience factor.

Simplex/Duplex

Modems may be designed to operate in three modes: (1) Simplex, where data is transmitted (and hopefully received) in only one direction, (2) Half-duplex, where data can be transmitted in only one direction at a time, but that direction can be reversed, and (3) Full-duplex, where data may be transmitted in both directions simultaneously. When referring to interactive terminals, half and full-duplex must be understood in the broader context of computer network conventions. Half-duplex refers to the case where the user's keyboard is connected to the printer at the terminal, and to the convention that when the computer is producing output it does not respond to input from the user (except perhaps for a special "break" signal). When the computer is not typing it is in a mode to accept input from the user. Full-duplex refers to the case where the user's keyboard is locally disconnected from the printer mechanism. When characters are received at the computer, they are returned to the user. The returned character or characters may not be identical to the received character; for example, some systems will return a control character as an up-arrow followed by the character ("ETX" is sometimes returned as "↑C"). The process of sending the received characters, exactly or as modified, is often referred to as the echo-plex mode of operation.

Full-duplex, however, also implies the separation of input and output to the point where a user at a terminal can be typing input while the computer is sending him output. In this case, "echoing" will be delayed until the output has been completed. It is not unusual, nor particularly distasteful, to hear the terms full-duplex and echo-plex used interchangeably although it is preferable to think of echo-plex as a subset of full-duplex (that subset related to the actual echoing of characters).

Multiplexing

One advantage of network communications lies in the cost advantage of high capacity communications links. But a mechanism must be provided to share these facilities among multiple users. Frequency division multiplexing (FDM) lets a number of low-speed devices share a link by dividing its frequency spectrum into a number of sub-channels. Time division multiplexing (TDM) assigns each low-speed device a time slice into which it may place information. TDM multiplexers operate on bit-interleaving and character-interleaving modes. The term multiplexing is generally applied to non-programmable devices.

Polling

Polling is another technique for controlling the use of lines by an agreed protocol between devices trying to share a common transmission path. The devices are rigidly controlled (so that only one of them sends information along a line at any instant) by an exchange of control signals or messages between them. Sometimes polling is governed by the central computer which sends a control message to each terminal in turn, inviting it to transmit an information message. The terminal replies either with such a message or with a control message indicating it has nothing to report. When the polling is done in this way great flexibility is obtained, because the computer may reorganize the order of polling should this become necessary. But rather a high proportion of control messages have to pass through the network and this tends to make it rather inefficient. Another technique of polling, possible when multiplexers are used, is to allow each remote multiplexer to poll the terminals connected to it. This is more efficient than polling from the central computer because each multiplexer is operating in parallel, and there are fewer control messages involving the computer itself. An alternative technique known as "hub-polling" is used on some multipoint circuits. With hub-polling, the computer invites the first terminal to send a message, if none is ready the first terminal passes the request to the second terminal and so on. This is particularly efficient when terminals are generally inactive and lines are very long. Whenever a terminal replies with an information message, the computer deals with it and then resumes the polling sequence by inviting the next terminal to proceed. In this way an active terminal near the beginning of the circuit is prevented from monopolizing the attention of the computer.

Network Processors

Simple multiplexing is implemented as an economy measure to increase the cost-effective utilization of communication facilities. More complicated algorithms and the declining cost of programmable processors have made it attractive to replace hardwired multiplexers with minicomputers. The inherent flexibility of these computers in turn made other functions possible. For example, by using small processors as the concentrators, the dialogue between these and the front end processor may be more complex. This can improve the efficiency of the network by permitting more sophisticated line control procedures, which may also facilitate recovery from failure situations. Again, some of the functions previously performed by the main computer, or the front end processor, may be delegated to the concentrator giving a more immediate response to the terminal. The polling of local lines, or resolution of contending terminals, the checking and formatting of messages, the correction of errors, guidance to operators and similar tasks, are typical of those conveniently performed by a concentrator with a resonable degree of processing power.

Front End Processors

In network communication many of the tasks performed are relatively simple but highly repetitive and can make considerable demands on the time of a computer. It oftens proves more economic to perform these functions by a small separate computer called a front end processor. The main advantages of this approach are:

1. The cost of hardware to attach lines is often less with a small computer, possibly because different constructional methods are used.
2. The processing load removed from the main computer will considerably increase the power available for computational purposes.
3. It becomes possible to separate the complete system cleanly into two parts: the main processor and the communications network. This gives increased flexibility and may allow one part to be enhanced or replaced without affecting the other.

Concentrators

When synchronous (fixed time slot) TDM concepts are extended to adaptive synchronous behavior, a concentrator is required. (In this tutorial the term multiplexor is used to describe a non-programmable device, while concentrator refers to the programmable processor. The distinction is not precise and is becoming more vague as firmware enters the picture.) Concentration basically involves store-and-forward operation wherein a buffer in the concentrator is employed to accumulate chacacters coming from a terminal until a termination condition, such as buffer overflow or an end of message segment singal, is received. The accumulated message is then transmitted over the high speed network circuit, thus taking optimal advantage of the available capacity. Error control procedures and message reassembly may be implemented in the concentrator as well as terminal speed recognition and code conversion. The concentrator may reformat the data in order to interface the different speeds and protocols of the network and terminal circuits.

Switching Processors

Switching computers are also relevant to the discussion of communications hardware. In order for a user to establish a connection with a particular computer, a unique number can be defined to make that connection, the number being different for each computer to which he wishes to connect (the number can also be given to a computer which will dial it by using an automatic calling unit). If a user, however, is to dial only one number, or be connected to one leased line regardless of the computer or computers he wishes to communicate with, then a switching computer must be employed in the network architecture. The switching computer performs the store-and-forward function; that is, a message is received in route, stored only until the proper outgoing line is available, and then retransmitted. In one technique, each message contains an identification of the sending device and an address for the receiving process allowing the switching computer to route messages from any source to any destination. The switching function can also be performed by mapping the input ports of the switching computer to the appropriate output ports. Besides the inherent flexibility in software switching, the store-and-forward technology allows for optimal line utilization, since interprocess messages can be mixed with interprocess messages involving different host computers and terminal devices on the network. Also, the switching computer can perform other functions such as concentrating terminals and interfacing to service computers.

Error Detection and Treatment

The purpose of error control is to ensure that the information received by an acceptor is as intended by the source. There must be some means for the acceptor to deduce when received information contains errors, and a mechanism for removing them. Two types of error control may be identified: forward error control where sufficient redundant material is included with the information to allow the acceptor to detect an error and to infer the correct message, and feedback error control, where some redundancy is needed to reveal errors, but correction is made by retransmission. The introduction of redundancy decreases the capacity of the channel for carrying information. The choice of which scheme to employ for a particular application depends upon the ease of introducing the redundancy, the kinds of error possible with a given type of link and the performance criteria to be met. Published error rates of 1 in 50,000 for 50 baud lines to 1 in 500,000 for 600 baud lines give a false impression of a random variable. Switching, impulses, crosstalk, central office problems and maintenance personnel cause transmission lines to exhibit time varying burst error characteristics.

Software

This section will concentrate on network communications software and the identification of components in which the communications software resides.

Minicomputers as Communication Processors

There is a significant trend in the application of minicomputers as communication processors due to their relatively low cost and flexibility. Minicomputers have been applied as front-end processors where they connect between a central computing facility and a communications network, performing communications control functions for the central computer. As remote concentrators, minicomputers are located close to terminals and allow many terminals to share a single high speed circuit connecting to the central computer facility. As device controllers, minicomputers can control the high speed readers and printers of the central computing facility, relieving the central computer of that task; or they can control high speed readers and printers at locations remote from the central facility, permitting those devices to be placed when they are more needed. Minicomputers can also be message switching computers in large networks supporting the overall connectivity of the network and routing message from multiple sources to the correct destinations.

Front-End Processors

Front-end processors for a computing complex are often employed as device drivers, controlling the peripheral devices of the central computer, and as communication controllers handling all messages between terminals and programs in the central computer. In performance of these tasks, the front-end processor is responsible for the following:

1. Line control of all communication circuits between the central computer and remote terminals
2. The control of the interface, typically through a memory channel, to the central computer
3. Message formatting to translate messages to a format compatible with the destination computer or device
4. Error checking and correcting to guarantee the integrity of transmission
5. Device multiplexing to assure the efficient use of channels and circuits
6. Device control allowing for flexibility in the connection of different peripherals to the central facility.

The major advantages of front-end processors are in offloading the central computer processor, memory, and storage facilities, and in providing flexibility in the number and variety of devices that can be connected.

Remote Concentrators

Remote concentrators allow terminals at remote locations to share communications circuits. Due to the rate structure of the common carrier, considerable savings can be achieved through the use of higher speed circuits rather than a larger number of lower speed circuits with the combined equivalent bandwidth. In addition, in contrast with fixed-slot time division multiplexers, concentrators allow for the statistical multiplexing of the communication circuit such that bandwidth, unused by some devices, can be given to devices that require it. Remote concentrators are responsible for terminal handling, sometimes for peripheral device control, buffer management, and error control between terminals or devices and the control computing facility.

Network Communication Processors

Often communication processors in a network situation perform multiple functions. For example the ARPA Network Terminal Interface Message Processor (TIP) acts as a remote concentrator and a message switching computer. The TYMNET Base Tymsat performs as a front-end processor and a message switching computer. For the purpose of this discussion, these types of devices can be considered as network communication processors.

Communication Control Procedures

Communication control procedures are conventions for controlling the flow of messages between communicating processes. In a network situation, communication control procedures can perform the following:

1. Identify the source and destination stations
2. Request permission to send
3. Invite a station to send
4. Arbitrate contention when multiple stations or devices require the same communication channel
5. Provide for error recovery
6. Identify a device or process
7. Identify a message when multiple messages are outstanding between communicating processes.

Phases of Control

The phases of communication typically involve establishing the sender and the receiver, transferring the message, terminating the message transfer, and terminating the connection.

Error Control

In order to insure that transmission is occurring without errors being introduced into the data, it is necessary to establish a set of conventions between sender and receiver for detecting and correcting errors. The simplest form is a parity check on each transmitted character. This is often called a vertical redundancy check and is used to provide protection against single bit errors within characters. The longitudinal redundancy check (LRC) provides for a check across an entire message. This is done by computing a parity bit for each bit position of all of the characters in the message. The most powerful form of check is the cycle redundancy check where polynomial division is applied to the message string and the remainder used as the check digit(s). The same computation is performed at the receiving end, and the results compared. When a receiving station receives a message and determines that it is error free (using some combination of the above checks) it informs the sending station by transmitting an acknowledgement of the message (ACK). If a receiving station detects an error, it may transmit to the sending station a negative acknowledgement (NAK). In case the positive or negative acknowledgement is itself lost in transmission, sending stations typically have a time-out period, after which, if no acknowledgement is received, the message is re-transmitted. Many systems in use today rely on the acknowledgement (ACK) and time-out convention and do not employ the negative acknowledgement.

Common Heading Information

Messages are composed of data and control. The control information is contained in that portion of the message called the heading. Some typical heading information is the following:

- Identification of the originating station
- Identification of the sending and receiving device or process
- The priority of the message
- The security class of the message
- Routing information concerning the distribution of the message once it has reached its destination
- Message processing information concerning its status as data or control.

Protocols

Thus far, the discussion has been confined to the basic communication control procedures used for controlling a communication link between two processes. In large, complex networks it is desirable to build upon these basic procedures and form higher level protocols to control the flow of messages between processes, distributed among the many computers and devices which may be interconnected by the network. A protocol is essentially a set of conventions between communicating processes on the format and content of messages to be exchanged. To make implementation and usage more convenient, in sophisticated networks higher level protocols may use lower level protocols in a layered fashion.

Standardization

An industry-wide, internation design effort has begun to create a set of standards to support open system interconnection (distributed systems).

The era of distributed information systems crept into the computer field many years ago. From the moment the first communication line was attached to a computer, we were into distributed systems. Time sharing brought a wider usage to the concept of distribution access. Remote batch systems introduced batch users to distributed systems. All of these systems can be categorized as "first generation" distributed systems.

The ARPANET project, begun in 1968, was clearly recognizable step into the second generation of distributed system. It was an attempt (successful) to allow the users at their terminals to have access to a number of different computing facilities which were distributed across the United States. It offered many capabilities to the user, such as file transfer between computers, and remote time sharing and remote batch processing. This time, distributed I/O was combined with distributed processing. A new concept, it received most of its public attention for its packet switched network capability. This very important feature was a means for sharing communication lines in the network, on a demand basis. It provided the distributed information world with the flexibility that the automobile and truck have given to our transportation system. It permitted data to be sent to an arbitrary destination, when required, by whatever route was available. The data stream was broken up into small elements (packets), each of which had sufficient labelling to enter the network and be permitted to travel, on its own cognizance, until it reached its destination. There the packets would be reassembled into the original data stream. The communication lines were used just as our modern highways, on a first-come-first-serve basis, with alternate routes as detours, when the primary route was blocked or busy.

The work of CCITT, publishing the X.25 protocols, provides the basis in standards for national and international networks which can be used to connect terminals and computers using the packet switched network technology.

More recently, a number of computer manufacturing companies have been announcing their distributed systems products, i.e., Burroughs (Burroughs Communication Architecture), DEC (DECNET), Honeywell (DSE), IBM (SNA). Each of these announcements has told how it would be easier to interconnect the computer and terminals of that manufacturer and support distributed systems within their product lines.

In late 1976, a major effort to provide higher level interconnection across diverse computer lines and international boundaries was initiated by the British Standards Institute (BSI). They quite clearly saw that it would be impossible to get all the computer companies and all the telephone companies going in the same direction without an international standards effort to coordinate the activity. They could also see that the current individual effort of the vendors would lead to imcompatabilities which might take years to straighten out, if something was not done quickly.

The goal of Open System Interconnection is to define standard interface and standard protocols which would permit any person, any terminal, any physical process, and any computer which used these standards to exchange information. It would include well-known computer functions as remote job entry, time sharing, administrative message switching and file transfer.

Even though the benefits possible from such standards are incontrovertible, the task in establishing such standards is enormous. The organizational problem alone is incredible. The technical problem is bigger than any one previously faced in information systems. And the political problems will challenge the most astute statesmen.

In response to the BSI initiative, a new subcommittee was established "Open System Interconnection."

In August, 1977, the ANSI/X3/SPARC/Study Group - Distributed

Systems began meeting in Washington, D.C. with J. S. Foley (Burroughs) as the convener. Mr. Foley sits as a member of SPARC and served as the chairman of the team which wrote the study group's program of work. The key point of this program was to create a "reference model" by which the various elements of the distributed system could be organized and studied. As a study group, its task was to study and recommend to SPARC what standardization projects should be considered.

One of the interesting aspects of the ISO Sub-Committee 16 - Open System Interconnection was the fact that its interests overlapped the previously assigned area of responsibility of the ISO Sub-Committee 6 - Communications. This ambiguity was mapped directly back into the various national bodies where there existed a technical committee, national analogues to SC6. For example, within ANSI, there existed X3S3 - Communications. This ambiguity has been resolved in two ways within the U.S. groups. First of all, by the fact that X3S3 was a ANSI technical committee charged with preparing standards specifications. In contrast, the study group was charged with developing an architecture which included within its scope the communication area. Second, the membership of the study group included within its members and consultants some of the most active members of X3S3 and SC6. G. W. White of the National Communications Systems was a member of both organizations. W. F. Emmons (IBM), serving on both X3S3 and SC6, acted as a consultant in getting the communications end of the reference model into focus. R. desJardins (NASA), J. P. McGovern (Univac) and S. R. Kimbleton (N.B.S.) also served as key members in getting the ANSI Study Group under way and beginning to establish the reference model which was the first target.

The work of the CCITT, with regard to public packet switched networks, also fell within the area of interest of SC16. CCITT had recently published its standard X.25 which defined the virtual call (virtual circuit) interface to a public packet switched network. Thus, again there were overlapping interests which needed to be coordinated, if a satisfactory set of standards were to be prepared which would facilitate the kind of open system which was envisioned.

During the early fall of 1977, the secretariat for SC16 was assigned to the United States. This made it possible to begin planning the first meeting of SC16. The date of February 27 to March 1, 1978, was set, with Washington, D.C. as the location.

The SC16 meeting subdivided itself into three ad hoc committees to examine specific areas. One ad hoc committee, lead by H. Zimmermann, chief of the French delegation, studied the problems of "the overall architecture." They studied the criteria for the model control vs. data interfaces, justification for levels, formalization of definition, and terminology. Another ad hoc committee, lead by C. Solomonides of the U.K., concentrated on the first four layers of the model. This group was categorized as making up the "transport service." The four layers, from the bottom up, were named "physical control," "link control," "network control" and "transport end-to-end control." Their interest was in clarifying the description of these layers and in describing the interface between the Transport Service and the "users" of the transport service. The third ad hoc committee, lead by R. desJardins chief of the U.S. delegation, concentrated on the top three layers of model. This group was categorized as the "Users of Transport Service." These three levels, starting from the Transport Service, were named "session control," "presentation control" and the "application layer."

The reports of these groups were so successfully received that it was decided to create three working groups within SC16 to carry on the development work between meetings of SC16. The three working groups were patterned after the ad hoc committees and charged with continuing the work begun in Washington, D.C. The working groups established and their conveners are:

. WG1: Architecture - Open System Interconnection, H. Zimmermann (France)

. WG2: Users of Transport Service - Open System Interconnection, A. Langsford (U.K.)

- WG3: Transport Service - Open System Interconnection, G. W. White (U.S.A.)

One of the significant challenges, proposed by the U.S. presentation to the first meeting of SC16, was that an initial set of new standards be produced on a very short schedule. These included:

- Transport End-to-End Control protocol and interface by August, 1979
- Session Control protocol and interface by December, 1980
- COBOL CTG version of Presentation Control by 1980
- Remote file access protocol by December 1980, and
- Remote job entry protocol by December 1980.

The December, 1980 date for the initial standards is based upon three factors. First, that there are existing technical solutions for the problems. Second, that it is critical to get these standards into existance before too many ad hoc or one company solutions are installed. Third, that the effective life time of a member of such standards' activities is about three years.

Possibly the most significant insight was one given recently by Prof. Maurice Wilkes of Cambridge University (U.K.). He said, "You people are not doing classical standards work, you are doing an international, industry-wide, design effort under the guise of standardization!"

CONCLUSION

From this short "tutorial" it is clear that, as this study reveals, the complexity of networking is such that the people involved must be prepared to provide certain essential information such as:

1. Number and locations of terminals desired
2. Data base size and content
3. Files to be accessed and interrelated
4. Response times required for efficient use
5. Error detection required
6. Number and type of applications to be run on the system
7. Funding available

before any design of consequence can be done.

APPENDIX D: DATA BASE MANAGEMENT SOFTWARE

To begin with, clarification must be given as to what is meant by "data base management systems" as opposed to a similarly named, but distinctly different, class of software often called "data management systems."

A "data base management system" can be defined as a software system that is intended to manage and maintain data in a non-redundant structure for the purpose of being processed by multiple applications. A "data base management system" organizes data elements in some predefined structure, and retains relationships between different data elements within the data base.

A "data management system," on the other hand, is one that is intended primarily to permit access to and retrieval from already existing files (usually for a single application). Although a "data management system" may provide the capability to minimize data redundancy and centralize the storage of data, the principal intent of the system is to perform such functions as information retrieval, report generation and inquiry for a single application.

Since, in this study, we are addressing ourselves to a comprehensive data base capable of containig files to accommodate multiple applications, our interest is in "data base management systems."

The choice of the data base management system by the individuals governing the network should be of utmost importance. Nearly all computer manufacturers offer some sort of data base management system. One must be careful, however, to ascertain if the system is transportable; that is, whether it will run on another manufactuter's machine. Most data base management systems are designed to be independent of the operating system but, in fact, most systems are sensitive to logic changes in any operating system version.

The type of system needed is determined, in large part, by the data base to be manipulated or managed. If the network is to handle full MARC records and allow for reference retrieval as opposed to simply dealing with circulation records or a regional sub-set of the data base, a more powerful data base management system will be required.

It is unfortunate, but as of this date, there has been no data base management system developed that lends itself specifically to MARC records. Most systems on the market today are generalized. One exception to this could be FLIRT, the Data Base Management System developed by CLSI. Its drawbacks, however, are that it is imbedded in DEC's RSX 11-M operating system and, therefore, will only run on a DEC machine.

For this study, several data base management systems were explored. It was determined that a generalized data base management system should incorporate most, and preferably all of the following features:

- Application program independence from the Data Base Management System (DBMS) control programs
- Support of the programming language used in the environment before installation
- Utility programs to facilitate creation and maintenance of the data base(s)
- Facilities for data reorganization
- The ability to affect data security and access limitation
- Automatic re-start capabilities in case of system failure or the ability to recover operations manually with minimum effort
- System facilities for "fine tuning" of the DATA BASE MANAGEMENT SYSTEM (DBMS)

Of the systems studied, three seemed to contain most or all of the features mentioned. The three systems are:

1. ADABAS - Although the most expensive, this system is by far the leader in data base system software. However, unless the network expands the data base to become extremely large, it could be overkill. Of interest, however, DEC has announced support of a subset of ADABAS at a cost of $70,000 per CPU.
2. IMS - There are a number of IBM 360, 370's in the Commonwealth.
3. TOTAL - This system has the largest group of users (or purchasers) and is one of the few that runs on main frame and some mini-computers.

The ideal Data Base management and/or operating system should have the following features in order to handle a data base of full bibliographic records.

1. Multi-access from both on-line terminals (local and/or remote) and other computer systems. Protection should be provided against inadvertant and unauthorized access and/or update.

2. Variable format data records with variable length fields and sub-fields for efficient use of storage space.

3. Dictionary (Directory) driven to provide efficient access to data records.

4. Each data record should have a unique key assigned by the Data Base Operating System.

5. All records should be defined by a directory entry describing its fields, sub-fields, keys, and other characteristics (i.e. type, edits, validation, size, decimal, etc.)

6. All data records should be indexed both statistically (at creation time) and/or dynamically (as needed). Keys should be deleted, changed, and/or added without affecting the data records.

7. The data base structure should allow the creation of a sub-set of the data base without duplication of the data records.

8. A Data Base query and/or selection process should be available for both on-line and/or batch usage to create sub-sets of the data base. These functions would make use of the record descriptor (5).

9. A facility to allow gross changes across the data base and/or sub-sets of the data base. These functions would make use of the record descriptor (5).

10. A facility to do stand-alone counting of records and tallying of fields in the records of the data base and/or sub-sets of the data base. These functions would make use of the record descriptor (5).

11. A report formatting facility should be available for the data base and/or sub-sets of the data base.

12. System modules should be provided to perform the following functions:

    a. Request fields
    b. Display fields
    c. Change fields
    d. Validate fields

13. Sort/Merge capability to create a sub-set from two or more sub-sets. These functions should be available both as on-line (interactive) and/or batch. This function would make use of the record descriptor (5).

14. Transformation capability to change internal formats to external format (and vice versa) to allow interchange of data between systems.

15. Store-and-Forward message capabilities, both informational and operational request.

If the system contains these features and they can be demonstrated to the potential user, then it is truly "state-of-the-art" and has built into it the ability to adapt easily to changes in both software and hardware technology. ADABAS is the closest, but even it does not meet all the criteria.

ADABAS: (Adaptive Data Base System) is an interesting and powerful data base management system that competes with IBM's IMS, Cincoms TOTAL, and several other systems.

The system was released in 1970 and became available in the United States in 1972. It is the Data Base Management System used by the Washington Library system.

While ADABAS ranks as a top-notch data management system, few of its features offer unique capabilities. Rather, the system has been put together with such thoroughness and skill that its most impressive features are not directly evident to the end user and appear in the form of reduced system overhead.

For instance, ADABAS uses a data compression algorithm to load data into the data base. This simple technique is widely available on a stand-alone basis, but its incorporation into ADABAS as an integral function of the system typically results in about 30 percent reduction in the total volume of the structured data base as compared with the original size of the raw, unformatted data before being loaded. This facility alone would aid in the manipulation of the unwieldy full MARC record.

Another unique feature of ADABAS is the separation of physical data storage from the representation of logical relationships of the data base. This concept permits the representation of completely general network structures and eliminates the necessity for data base reorganization - two characteristics not exhibited when hierarchial and chaining techniques are employed as in several other data base systems.

Although ADABAS is by far the most efficient DBMS on the commercial market today, it may be overkill in the case of a Library Network for the Commonwealth. Network members should keep in mind what role the data base will be playing in the future. If it is decided that Virginia will build a data base of holdings within the Commonwealth and eventually add other states' holdings to the base and allow their use of the data base and application programs, then consideration of this type of data handling could be in order. The logic and techniques inherent in ADABAS are also present in some other DBMS systems which have not been put on the commercial market. Duplication of these could be considered as a viable alternative to purchasing ADABAS.

ADABAS - Characteristics

- SUPPLIER: Software AG of North America, Inc.; 1180 Sunrise Valley Drive; Reston, Virginia 22091. Telephone (703) 860-5050
- BASIC FUNCTION: DBMS with a number of utility programs for use under DOS, OS, or their VS counterparts, with BDAM used for data base generation and access.
- ADABAS operates as a 'HOST' language system with high-level CALLS supported by ADAMENT, the ADABAS macro interface.

Also available is ADASCRIPT, a self-contained natural language query operation which allows for full Boolean search of the data base.

. OPERATION: Applications are defined in terms of functional requirements and types of data to be processed.

Individual application programs may be written in COBOL, PL/1, FORTRAN, or Assembler language. Calls are inserted at appropriate points for the ADABAS access commands.

ADABAS has internal functions to handle data requirements unlike a system such as IMS which requires a Data Base Administrator to handle complex data base design.

ADABAS supplies interfaces for most of the commercially available communication monitors.

Retrieval timing example: Five-element set of search criteria against a one-million record data base with 200 fields per record took four seconds to determine the number of qualifying records and read the first record. The performance test was run on an IBM 360/50.

. HARDWARE/SOFTWARE REQUIREMENTS: ADABAS consists of a 130K-byte NON-OVERLAID nucleus with about a 1-million byte disk-resident library including the utilities. The nucleus (130K) plus a variable-sized buffer area (minimum 30K) must be resident in main memory at all times. In a multi-user environment, a re-entrant 8K to 16K front-end module is used to queue access requests. ADABAS can be run in a 170K partition or region under DOS, OS, or their VS counterparts, on an IBM system/360 or 370, or on an IBM-like UNIVAC 9000 series or Siemens 4004 system. The system also runs on an IBM-like AMDAHL and has recently become available on a DEC 20.

. PRICING: Provided on a permanent license basis for $120,000 for the first CPU, including all normal training. The second CPU license costs $60,000. Lease plans range from month-to-month at $5,000 per month to five years at $2,500 per month plus $2,000 training costs.

. CURRENT USERS: Over 130 installations with more than 75 in North America.

TOTAL: is a host-language data base management system implemented much along the lines of the CODASYL Data Test Group Report, except that other host languages in addition to COBOL can be used.

TOTAL performs in both batch and on-line environments using their own data communications module (ENVIRON/1). SOCRATES is available for data retrieval and display.

As compared to IBM's IMS, TOTAL provides a more streamlined approach to data base management.

Basically, TOTAL is oriented toward simplifying application programming problems by handling complex relationships among the various data items that constitute the information base. In the case of a typical library system, these would include the bibliographic file, patron file, vendor file, invoice file, circulation file, etc.

One of the nice conceptual features of TOTAL is that a user may start small, with perhaps a single application (circulation for instance) using two or three data files, and grow in modular fashion to very large and sophisticated data base/data communications structures with very little impact on the previously implemented systems.

TOTAL, unlike IMS, allows the user to utilize any logical data structure (network or hierarchical) in conjunction with a single physical data organization.

TOTAL provides a simple and straightforward set of facilities for organizing a data base and for manipulating the data in the data base within an application program. TOTAL supports any host language that contains a CALL statement (COBOL, FORTRAN, PL/1, Assembler), therefore all data base access is done through CALL statements.

TOTAL's greatest limitation (unlike ADABASE) lies in the fact that the defined data structure is fixed and not dynamic once the data base is loaded. Records can be added to and deleted from any of the existing data sets; but if new relationships (linkage paths) are to be defined, new data sets cannot be added and disc storage areas cannot be expanded without at least partial regeneration of the data base, which may mean the reloading of all affected data files.

TOTAL, however, is one of the most effective high-powered software systems in use today and has been successful in replacing IBM's IMS in numerous installations.

TOTAL - Characteristics

. SUPPLIER: Cincom Systems, Inc.,
2300 Montana Ave., Cincinnati, Ohio 25211.
Telephone (513) 662-2300

. BASIC FUNCTION: TOTAL provides facilities for generation of a data base that permits automatic cross-referencing among data records. A facility is also provided for accessing the data base from conventional application programs written in COBOL, FORTRAN, PL/1, Assembler, etc.

- OPERATION: The basic system includes three phases:
  - generating the program for controlling the data base structure
  - pre-formatting the disk areas
  - controlling the access to the data base.

  All phases of TOTAL are loaded at execution time which speeds up considerably the loading of the data base.

  TOTAL allows for multi-tasking and for batch processing. A given application can process the entire data base or any portion desired by defining a subset of files and requesting that definition by name at execution time.

- HARDWARE/SOFTWARE REQUIREMENTS: TOTAL is currently operational on IBM system/360 and 370 computers running under DOS, DOS/VS, OS, VSI and VS2; Honeywell series 200 and 2000 computers running under MOD 1 (MSR), MOD 2 and OS/2000; UNIVAC series 70 and 9400/9700 running under TDOS and DOS; NCR Century Series; IBM System/3 and CDC Cyber Series. A version of TOTAL is also available on VARIAN (now UNIVAC) V70 minicomputers. Host languages include COBOL, FORTRAN, PL/1, RPG II and Assembler language.

  Main memory requirements range from a minimum of 3K bytes for the read only version to 30K bytes for the full "CENTRAL" Version (allows both batch and multitasking). These figures exclude application code and I/O buffers.

- PRICING: Prices vary depending upon the computer to be used. For IBM DOS and DOS/VS installations the base purchase price is $34,500 and the rental price is $825 per month. For IBM OS and OS/VS, and corresponding CICS installations, the base figures are $39,500 to purchase and $1,150 per month to rent.

- CURRENT USERS: Over 1000

  NOTE: Cincom Corporation has announced just recently that they are offering a natural language query module to work in conjunction with TOTAL. Information is sketchy; however, it is supposed to allow full Boolean search capability. Its cost in conjunction with TOTAL, ranges from $120,000-150,000.

IMS: IBM's Information Management System is one of the most controversial and frequently discussed software systems on the market today. It represents the criterion by which other vendors judge or compare their data base management systems, and it has introduced numerous data management concepts that have been copied, and frequently improved upon, by other vendors.

IMS is one of the most flexible systems on the market today; however, flexibility often carries with it the complexity of creation and implementation, not to mention use.

IMS is not the answer for all types of user environments and if users find that it does not meet their needs they will be hard pressed to convert to another system without incurring a great deal of expense and lost time.

The fundamental functional concept of IMS for data base management is fairly simple. Utility programs are used to describe the structure of the system from two view points: stored data structure as seen by the system and logical data structure as seen by an application. Stored data is described only once, but many descriptions of the logical data can exist.

IMS is a "host language" system, the user writes application programs in COBOL, PL/1 or Assembler language for which interface modules are provided.

The use of IMS requires a data base administrator. It is this person's job to make certain that all structures and paths within the data base are correct.

IMS could be used under certain limited applications, however, it would certainly not be the most effective or efficient for a total comprehensive library networking system.

IMS - Characteristics

. SUPPLIER: IBM Corporation, 1133 Westchester Ave., While Plains, N.Y. 10604. Telephone (914) 969-1900.

. BASIC FUNCTIONS: IBM provides the capabilities for generating and accessing a data base, with automatic cross-referencing among data records. IMS/VS operates under OS/MFT or OS/MVT.

Both IMS/VS and IMS-2 offer on-line message processing with an optional data communication feature.

. OPERATION: The user must perform the following steps in order to use IMS:

.. Applications must be defined in terms of functional requirements and types of data to be processed. Individual application programs must be written in COBOL, PL/1 or Assembler language to perform these applications.

.. All of the individual data requirements for each application must be coordinated into an overall data base contents requirement. This is where the Data Base Administrator comes into play. Based upon the processing logic of various application programs, the administrator then selects one of the hierarchial IMS structures for the physical storage of data.

.. The administrator must then decide what type of communications and other on-line retrieval mechanisms are required.

.. Then the administrator must put together the IMS modules needed to support the installation. This task, in itself, can be most complicated.

Just a quick comparison between the operation for ADABAS and/or TOTAL with IMS makes it clear why many users are changing.

- HARDWARE/SOFTWARE REQUIREMENTS: IMS/VS or IMS-2 run on IBM hardware or IBM-like hardware only. All of the modules are written in IBM Assembler language.

- Core and storage requirements vary considerably depending upon the power built into the IMS system and its supporting communication and multi-tasking modules. To say the least, core and storage requirements are much greater then either TOTAL or ADABAS.

- PRICING: IMS-2, IMS/VS and all supporting programs and features are standard IBM Program Products. These are available from IBM on a monthly lease plan. Prices change regularly but an estimate for a data management system comparable to TOTAL would be approximately $3,000 per month.

- CURRENT USERS: IBM will not release statistics on the current number of users. However, Datapro estimates approximately 500 users of IMS/VS and IMS-2.

Conclusions

The choice of this particular piece of software for a comprehensive library network must be given the greatest thought and cannot be done until the following questions have been explored and thorough answers obtained:

1. Size and content of the base
2. Number of users and applications
3. Response time requirements

APPENDIX E: HARDWARE

Host Computers

In the development of a network, once the data base and its contents and size have been decided as well as how the data is to be handled, what applications will be involved now and in the future, and what member locations (terminals) and usage requirements established, the next step is to decide on the size and type of host processor (computer), location of communication processors and distributed processors. Although the decision regarding this hardware is best made after the system's requirements have been determined, often this due order has not been followed. Frequently, hardware and data handling software have been chosen beforehand and the new data bases(s) and applications are squeezed into the existing setup. The result is not efficient.

It is to be hoped that Virginia will heed what has occurred elsewhere. In Ohio, OCLC has had to go through many trial and error periods because the system was first put up on existing hardware. It is only now that they have purchased hardware and software which will efficiently handle their problem. BALLOTS and University of Chicago have the same problem. In Washington, WLN was able to obtain, early on, the proper hardware and software to effectively implement their system. The processor is shared by others within the state, but since WLN is a "not for profit" corporation, their time is purchased just as it would be from a commercial vendor and therefore it does not run the risk of being "pushed off" the system in case it is deemed, by the legislative powers that be, that the computer is needed for some other purpose.

The type of processor (main frame, mini, micro) will depend upon the volume of work which must be processed and the expected response time to the user.

Since processors (computers) and storage devices (discs, tapes) are the greatest initial capital outlays, a great deal of study should take place among the prospective users of the network regarding the short and long range plans for the network and whether the hardware used to begin the network will be capable of expansion and modification as new developments occur. With regard to planning for new development, technology is expanding so rapidly that the ordinary lay person would wonder if any equipment purchased today would not be outmoded within a few years. Reading the claims of the various hardware vendors one might surmise this to be true, but in actuality it is not. As a general rule, modifications can be made to existing hardware to bring it up to the latest specifications. More concern should be given to the kinds of equipment to be used. Some hardware manufacturers supply interfaces which allow one type of computer to communicate with another. If this type of interface is not available, it could be quite costly to develop it. This should be carefully considered especially if the network begins with regional systems which may eventually interact. This could be best illustrated by what has happened in Virginia. There are four regions contemplating four different circulation systems:

1. Tidewater/Old Dominion. It is not clear what computers the Tidewater area will be using, but since their plans are to communicate with Old Dominion, which is planning on using Data General machinery, the computers installed in Tidewater must interface with Data General if the area hopes to have a true on-line base of combined holdings.
2. Henrico County plans to use UNIVAC processors for its circulation system.
3. VPI has a circulation system currently operating on a Hewlett Packard 3000.
4. Fairfax County has just opened bids for a circulation system. The low bidder was CLSI which uses a DEC mini.

To connect these various systems into a network to produce a data base of common holdings or even inter-library loan capability could prove to be very costly. The content, in fact even the layout, of the data bases is different. Software programs would have to be written to accomplish any communication between the various bases. Add-on application programs for acquisitions, cataloging, etc. could also present a problem. A program written for one system, even if written in an ANSI language such as COBOL, may take a great deal of modification to run on the varying systems.

If, however, the network is developed with these factors in mind, then accommodations can be made during the inception to allow communication, within a central communications processor, between these systems.

When evaluating various hardware vendors the members of the network should keep certain considerations in mind:

- Service - is service to the machinery easily available and readily done?
- Software - does the vendor offer the major software application languages (COBOL, FORTRAN, PL/1, RPGII, etc.) and if so, at what level of development?

  Does the vendor have a good reputation in supporting the software provided? Is it updated regularly, is it corrected if found faulty? What does the vendor charge for these languages?

  Does the vendor offer normal communication software (i.e. can a peripheral be connected to the central processor many miles away by direct line)? If available, how standard is it? Can it easily be interfaced with other vendors' software?

  Is a data base management system offered? If so, is it truly a data base management system or just a management or information retrieval system. What does it cost? Can it be interfaced with other systems?

  Is a free standing inquiry system offered? Does it do full Boolean arithmetic?
- Record Retrieval - As the data base grows and as more users are added what kind of degredation can be expected? How much core must be added to stop the degredation. At what cost?
- Peripheral Hardware - Is it serviced by the vendor? Is it standard, i.e., if a CRT or printer, can it be used on different computer systems?
- Future Developments - What has been the vendors past track record? A leader in the field of new technology, or a recent entry?

  If a leader, have new developments made prior systems obsolete or have provisions been made to easily upgrade these systems?

One might think that all vendors, especially in the highly competitive computer hardware/software business, would have addressed these problems and solved them. But that is not the case. The network members should make certain that an impartial evaluation of hardware and software be done once they have defined the network and its uses.

APPENDIX F: TURN-KEY SYSTEMS

Automated library systems which assist operations in the public services area (particularly, circulation) and which are able to serve large systems with an on-line capability are available from a number of vendors. All are designed to handle the basic circulation functions and to be flexible enough for integration or interface with one or more sub-systems of library operations (e.g., cataloging, acquisitions). Our intention here is not to examine the variations and differences among vendors with regard to applications, but to review varying technical attributes significant to networking. The list of vendors surveyed is not exhaustive but includes all who replied to our request for information--and this includes all major vendors in the field. Some information on costs has been given, but a strong warning needs to be issued against interpreting this as anything more than a very rudimentary estimate. Estimates approaching any degree of exactness would require knowledge of a complex of variables: number of library locations and their distance from the central location, number of checkpoint stations, number and kind of terminals for each (printers, CRTs, etc.), number of applications involved, number of titles, number of items, number of authors, amount of information per record, percent of records already in machine readable form and the format used, number of patron registrations, population served, amount and rate of circulation (volumes per hour at peak periods), projected rate of growth both for collection and for circulation, communication systems to be used, and availability of a host computer. None of these variables are at present fixed. Where costs were obtained from a vendor, these have been mentioned in the terms considered appropriate by the vendor (and always with the due caution stated above).

DATA PHASE

Computer: Data General Nova and Eclipse. Also can be run on DEC PDP 11 (complete line: 1134-1170) and IBM Series 1. Potential for use on DG Micro Nova

Programming language: Form of MUMPS

Management System: Uses a proprietary version of MMIS.

Storage: Capable of 2.4 billion bytes

Peripherals: Data General

Terminal support: With larger mini-computers, upwards of 1000 terminals.

Cost: For a system with 16 terminals, 192 million Bytes of storage, 64K of main memory, $200,000-250,000

CINCINNATI ELECTRONICS

Computer: Univac V-77 (200,400,600)

Language: FORTRAN

Management system: TOTAL

Peripherals: Univac printer, MI^2 printer, Data Products, Centronics

Storage: Up to 2,000,000 bytes for main storage. Disk: 4.8 billion bytes.

Terminal support: Up to 256 terminals. Basic system: 64 terminals.

Cost: Minimum system--$100-125,000. Up to 1.5 million dollars. Typical system with 8-10 branches: $250-300,000.

SYSTEMS CONTROL, INC.

Computer: Digital Equipment 1170 and 1134

Language: BASIC + 2 (timesharing)

Management system: Their own

Terminals: Universal Transaction terminals, Lear Siegler,
Printers: Digitex, Innovex

Storage: Main--512K bytes

Terminal support: 128 terminal maximum

Costs: For 32 terminal system, $500,000--80 terminal system--$750,000-1,000,000.

DECICOM

Computer: General Automation 16 series (20,440,550)

Language: FORTRAN

Management system: File Management System and their own.

Peripherals: Input terminals are designed by Decicom. CRT is bought from outside. System compatible with a number of standard CRT's.

Storage: Up to 2 billion characters.

Terminals: 256 maximum

Costs: For a library with five input stations and circulation of 2,000,000, $100,000.

GAYLORD

Computer: DEC 1170 and 1150 series

Language: Assembler, some host programs in FORTRAN

Terminals: Informer, Inter-Mec light pen, Printer: Printronics is being developed for bar-code products.

Storage: Virtually unlimited for DEC 1134.

Terminal support: Up to 100 terminals.

ICC/PLESSEY

Computer: Interdata 732

Language: COBOL-66

Management system: Proprietary to Plessey

Terminals: Circulation terminal specially designed by Plessey. CRT is optional and can be obtained from Lear Sigler ADM-3A.

Storage: 600 megabytes minimum, expandable to 2400 megabytes.

Terminal support: Up to 256 terminals

Costs: Minimum of $300,000 would support a ten branch library system)

CLSI

Computer: Digital Equipment 1134

Language: RSX 11-M

Management system: Their own

Terminals: Printer: Medium speed--Diablo High speed--Talley Printer (matrix style). Data Products Printer. Screen printer: SCI Corporation. CRT: Standard size--Lear Sigler, ADM model. Larger screen--Digilog

Storage: 1.2 billion characters

Terminal support: 1134 can support 64 terminals

Costs: Ranges from $38,000 to $280,000 dollars. For an average public library with one branch, $110,000.

UNIVERSAL LIBRARY SYSTEMS, LTD.

Computer: Digital Equipment (1134, 1160, 1170)

Language: BASIC + 2

Management system: Uses RSTS/E operating system

Peripherals: Digital equipment, Monarch light pen.

Storage: Maximum of 1.4 billion bytes

UNIVERSAL LIBRARY SYSTEMS, LTD. (Cont'd)

Costs: For a library with 140,000 circ. and a collection of 70,000, $125,000. For a library with 3 million circ. and a collection of 1,000,000, $400,000.

INNOVATED SYSTEMS, INC. (circulation system for the University of Texas at Dallas)

Computer: HP 21 MX

Language: FORTRAN with some Assembler for I/O

Management system: HP's IMAGE and Query systems

Storage: 7½ megabytes, expandable by disk. Runs HASP to IBM 370 for access to bibliographic information.

Terminal support: Up to 6

Peripherals: TI Silent 700 with dual cassette and TI printer

Costs: $150,000 (includes security and self-charge capability)

ACK - Positive acknowledgment character

Algorithm - A fixed step-by-step procedure for accomplishing a given result, usually a simplified procedure for solving a complex problem, also a full statement of a finite number of steps.

Analog - Pertaining to representation by means of continuously variable physical quantities.

ANSI - American National Standards Institute.

Application - System or problem to which a computer is applied. Reference is often made to an application as being either of the computational type, where arithmetic computations predominate, or of the data processing type, where data handling operations predominate.

ASCII - American National Standard Code for Information Interchange. The standard code, using a coded character set consisting of 7-bit coded characters, used for information interchange among data processing systems, data communications system, and associated equipment. The ASCII set consists of control characters and graphic characters.

Assembly language - A computer oriented language whose instructions are usually in one-to-one correspondence with computer instructions.

Asynchronous - Without regular time relationship; unexpected or unpredictable with respect to the execution of a program's instructions.

Attenuation - A decrease in magnitude of current, voltage, or power of a signal in transmission between points.

BASIC - Beginner's All-purpose Symbolic Instruction Code, a programming language with a small repertoire of commands and a simple syntax, primarily designed for numerical applications.

Batch - An accumulation of data to be processed.

Baud - A unit of signaling speed equal to the number of discrete conditions or signal events per second.

BDAM - Basic direct access method.

Bit - In pure binary numeration system, either of the digits 0 and 1. Synonymous with binary digit.

Boolean operation - A mathematical or logical operation that depends on the application of the rules of Boolean algebra. The operands and results assume either one of two values.

BPS - Bits per second. In serial transmission, the instantaneous bit speed with which a device or channel transmits a character.

Buffer area - An area of storage that is temporarily reserved for use in performing an input/output operation, into which data is read or from which data is written.

Byte - A binary character operated upon as a unit and usually shorter than a computer word.

Call - The action of bringing a computer program, a routine, or a subroutine into effect, usually by specifying the entry conditions and jumping to an entry point.

Card punch - A device that punches holes in a card to represent data.

Card reader - A device that reads or senses the holes in a punched card, transforming the data from hole patterns to electrical signals.

Centralized system - In network topology, simple time sharing system, with channels radiating from a central computer system.

Central processing unit - Principal unit of the computer that controls the processing routines, performs the arithmetic functions, and maintains a quickly accessible memory. Synonymous with main frame, processing unit and CPU.

Chaining - System of storing records in which each record belongs to a list or group of records, and has a linking field for tracing the chain.

CICS - Customer Information Control System.

Circuit - In data communication, a means of two-way communication between two data terminal installations.

COBOL - Common Business Oriented Language. An English-like programming language designed for business data processing applications.

CODASYL - Conference on Data Systems Languages.

Common Carrier - See Communication common carrier

Communication links - Physical means of connecting one location to another in order to transmit and receive information.

Communication common carrier - In USA, a government-regulated private company that furnishes the general public with telecommunication service facilities.

Communication software - Sets of software for monitoring communication activities include the following routines: Interrupt - upon a program interrupt, this routine directs data transfer between the communication control unit and the central processor and then returns control to the main program; Message Queuing - controls the order in which messages are stored, processed and transmitted; Error Control - corrects errors in messages received from other communication stations.

Composition - Network composition can be either heterogeneous or homogeneous depending on the similarity of the nodes or attached computers. A network can be considered heterogeneous if connected computer systems are of different types or manufacturers.

Concentrators - Any device that either combines multiple incoming messages into a single message or extracts individual messages from data sent in a single transmission sequence.

Control - Network control functions are distributed to various network components. The control functions include establishing the initial connection, flow control, routing, monitoring, maintenance and measurement.

Core storage - A magnetic storage in which the magnetic medium consists of magnetic cores.

CPU - See central processing unit.

Crosstalk - The unwanted energy transferred from one circuit, called the disturbing circuit, to another circuit, called the disturbed circuit.

CRT - Cathode ray tube.

Data base management - A systematic approach to storing, updating and retrieval of information stored as data items, usually in the form of records in a file, where many users, or even many remote installations, will use common data banks.

Data compaction - A series of techniques used for the reduction of space, bandwidth, cost, transmission, generating time and the storage of data. These techniques are designed for elimination of repetition, removal of irrelevancies and employment of special coding techniques.

Data compression - See data compaction.

Data set - The major unit of data storage and retrieval in the operating system, consisting of a collection of data in one of several prescribed arrangements and described by control information to which the system has access.

Data transmission line - A medium for transferring signals over a distance.

Disk storage (magnetic disk storage) - A magnetic storage in which data are stored by magnetic recording on the flat surfaces of one or more disks that rotate in use.

Distributed system - A type of network topology, where each node is connected to every other node. This may be a situation where the serving computer systems are geographically separated but serve a number of common terminals or terminal locations.

DLE - The data link escape character.

DOS - Disc operating system.

DOS/VS - Disc operating system/virtual storage.

ENQ - The inquiry character.

EOT - End-of-transmission character.

ETB - End-of-transmission-block character.

ETX - End-of-text character.

External Interfaces - Those interfaces at which host computer systems and a variety of terminals are connected.

FDM - Frequency division multiplex.

Flow Control - Control of traffic between the individual components will typically be controlled by ordinary line control procedures. Flow control between processes can involve special software in the computers on which these processes are running. An alternative is to use the regular software controller to control interprocess communication. Or have a control computer responsible for flow control accepting inputs from all processes in remote computers desiring to communicate.

FORTRAN - Formula translation. A programming language primarily used to express computer programs by arithmetic formulas.

FSK - Frequency shift key.

Hierarchy - A series of items classified by rank or order.

Host Computer - The central or controlling processing unit in a multiple processing unit configuration.

IMP - Interface Message Processors

Input/output - Pertaining to a device or channel that may be involved in an input process and at a different time in an output process.

Initial Connection - One of the network components must be responsible for establishing an initial connection between two processes desiring to communicate. This type of control can be centralized in the network such that every process desiring to communicate with another process in the network will do so by initially connecting to the same computer which will establish the desired connection.

Interface - A shared boundary. An interface might be a hardware component to link two devices or it might be a portion of storage or registers accessed by two or more computer programs.

Interleaving - Act of accessing two or more bytes of streams of data from distinct storage units simultaneously.

I/O - See input/output.

Keyboard - A systematic arrangement of keys by which a machine is operated or by which data is entered.

Linkage - In computer programming, the part of a computer program, in some cases a single instruction or an address, that passes control and parameters between separate portions of the computer program.

Load - To fill the internal storage of a computer with information from auxiliary or external storage.

LRC - Longitudinal redundancy check.

Mainframe - The central processor of the computer system. It contains main storage, arithmetic/logic unit, and special register groups. Synonymous with CPU and central processing unit.

Main storage - Program addressable storage from which instructions and other data can be loaded directly into registers for subsequent execution of processing.

Matrix printer - A printer in which each character is represented by a pattern of dots.

Measurement - It is possible to generate statistics to measure the internal performance of the communications network.

Mega - Ten to the sixth power, 1,000,00 in decimal notation.

Memory - See main storage.

Message switching - The process of routing messages by receiving, storing and forwarding complete messages within a data network.

Microcomputer - A complete general or special-purpose electronic digital computer, with CPU, storage, and I/O bus.

Minicomputer - A complete computer, built for on-site installation, that does not require a closely controlled environment. At one time used for specific tasks, such as manufacturing. Now used for all types of computing problems. Accepts a wide variety of peripherals, such as magnetic tape units, discs and line printers.

Mode - A method of operation; for example, the binary mode, the interpretive mode, the alphanumeric mode.

Modems - A device that converts data from a form which is compatible with data-processing equipment to a form that is compatible with transmission facilities and vice-versa. (from modulation and demodulation).

Monitoring and Maintenance System - A network system is monitored by control centers. Intelligent use of automatic remote checkout logic, including remote sensing of component failure, degradation and failure can be detected and personnel dispatched for repairs.

Multiplex - Process of transferring data from several storage devices operating at relatively low transfer rates to one storage device operating at a high transfer rate in such a manner that the high-speed device is not obliged to "wait" for the low speed devices.

Multiplexer - A specialized computer, with stored program capability, for handling input/output functions of a real-time system.

Multitasking - Multiprogramming that provides for the concurrent performance, or interleaved execution of two or more tasks.

NAK - Negative acknowledgment character.

Natural language - A language whose rules are based on current usage without being specifically prescribed.

Network Components - Computer networks are composed of a set of nodes and circuits.

Network control - In systems network architecture, a transmission control component that permits logically adjacent connection point managers to communicate through the common network, using sessions established for other purposes and thereby avoiding special session establishment.

Node - In a data network a point where one or more functional units interconnect data transmission lines.

Nucleus - That part of a control program that is resident in main storage.

Online - Pertaining to the operation of a functional unit that is under continual control of a computer.

OS - Operating system (full).

OS/MFT - The IBM System/360 Operating System that supports multiprogramming with a fixed number of tasks.

OS/MVT - The IBM System/360 Operating System that supports multiprogramming with a variable number of tasks.

OS/VS - operating system/virtual storage.

OS/VS1 - A virtual storage operating system that is an extension to OS/MFT.

OS/VS2 - A virtual storage operating system that is an extension to OS/MVT.

Peripheral equipment - In a data processing system, any equipment, distinct from the central processing unit, that may provide the system with outside communication or additional facilities.

Physical Circuit - Circuits which are dedicated on a hand-wired or leased basis, or are established by physically connecting the input path at each switching point.

PL/1 programming language - Designed to take advantage of recent developments in computer technology and to provide the programmer with a flexible problem-oriented language for programming problems that can best be solved using a combination of scientific and commercial computing techniques.

Polling - Interrogation of devices for purposes such as to avoid contention, to determine operational status, or to determine readiness to send or receive data.

Port - A functional unit of a node through which data may enter or leave a data network.

Printer - An output device that produces a durable record of data in the form of a sequence of discrete graphic characters belonging to a predetermined character set.

Processor - A device capable of receiving data, manipulating it, supplying results usually of an internally stored program.

Pulse train - A series of pulses having similar characteristics.

Reentrant - Attribute of a program or routine that allows the same copy of the program or routine to be used concurrently by two or more tasks.

Remote job entry (RJE) - Submission of a job through an input unit that has access to a computer through a data link.

Ring System - A type of network topology where each node is connected to exactly two other nodes.

Routing - The assignment of the communication path by which a message will reach its destination.

RPG II - Report program generator, version 2. A commercially oriented programming language specifically designed for writing application programs that meet common business data processing requirements.

Size - Refers to the number of nodes or number of connected computers or terminals.

Software - Computer programs, procedures, rules, and possibly associated documentation concerned with the operation of a data processing system.

SOH - Start-of-heading character.

Storage device - A functional unit into which data can be placed, in which they can be retained, and from which they can be retrieved.

STX - Start-of-text character.

SYN - Synchronous idle character.

Systems network architecture - The total description of the logical structure, formats, protocols and operational sequences for transmitting information units through the communication system.

TDOS - Tape disc operating system.

Terminal - A device or point where information can enter or leave a communication network.

Topology - Alternative network topologies or organizations include:

1. Centralized or star
2. Distributed network that is either partially or fully connected
3. A ring, variant of a distributed configuration; and
4. Various combinations of these.

Virtual circuit - In a data network operating in the packet switching mode, those facilities provided by the network for transferring data between the data stations that emulate those provided by a physical connection.

VS - Virtual storage.

INTEGRATED LIBRARY NETWORK
REQUIREMENTS AND RECOMMENDATIONS

AN ADDENDUM TO

INTEGRATED LIBRARY NETWORK SYSTEM

FOR

THE COMMONWEALTH OF VIRGINIA

A FEASIBILITY STUDY

TABLE OF CONTENTS

PAGE

Introduction 1

Network Components

Data Base 3

File Structures 4
Content 6
Relative File 7
Size 11
MARC Records 12
Accessing Criteria 15
File Updating 15
File Utilities 17
Security/Data Control 19
Processing 22
Conclusions 23

Application Software

Introduction 27
Cataloging Module 29

COM Catalog 31
Data Content 34
Retrospective Conversion Techniques 35
Network Schematic 44
Processing 44.1
Program Requirements 44.1
Hardware Requirements 48
Software Support Requirements 50

Data Base Management 50
Query Language 56
Communication 69
Operating System 69

Integration 70
Available Systems 70
Time/Cost Projections 71

Hardware 72
Software 77

Topological Configuration 80
Conclusions 84

Circulation 86

- Data Base Content 87
- Input Methods 99
- Program Requirements 99,105
- Support Software 100,106
- Output 100
- Hardware 100
- Integration 102
- Available Systems 102
- Time/Cost Projections 103
- Topological Configurations 106

Acquisitions 107

- Data Base Content 108
- Input Methods 112
- Program Requirements 114
- Support Software 116,117
- Integration 116
- Available Systems 116
- Hardware 117
- Inter-Library Loan 118
  - Files 118
  - Available Systems 120
  - Time/Cost Projections 121
- Serials Check-In/Control 122
- Reference Searching 122

Network Approaches 123

- Basic Steps 124
- Bibliographic 125
  - Cost Projections 126
- Regional 127
  - Cost Projections 129
- Combined 131

Network Administration 132

Attachment: An English Language Question Answering System For A Large Relational Data Base

INTEGRATED LIBRARY NETWORK REQUIREMENTS AND RECOMMENDATIONS

An addendum to "Integrated Library Network System for the Comwealth of Virginia - A Feasibility Study."

The study pertaining to the feasibility of establishing an integrated library network within the Commonwealth of Virginia determined that not only is it feasible to establish the network, but that such a network, if developed properly, would address the needs of the library community and their patrons in the following ways:

1. Relieving the Library Professional of most of the daily "busy work" so that his/her talents may be used in a more effective and efficient way;
2. Allow for easier exchange of materials between libraries;
3. Enable the library community to acquire material in a more efficient and cost effective way and by so doing, better serve the community;
4. Give the patron faster and more efficient response;
5. And many other less significant needs.

However, the finding which has greatest impact on determining hardware and software requirements for an integrated system within the Commonwealth is that the information needed to determine these factors is missing. (See Section 2 of Study)

The objective, therefore, of this addendum will be two-fold:

1. To give the data processing requirements for each component within a network - including hardware and software requirements for the component and time and cost projections;
2. To give a recommended modular approach to the eventual development of a fully integrated library network system.

It must be stressed that time and cost projections, data base size, processor and storage size are all dependent upon factors which have not yet been determined, as outlined in the study (Section 3). Therefore, in order to have some base from which to work, we will assume that the network will eventually be used by all members of the library community in one form or another and that the statistics used will be drawn from those gathered for the entire Commonwealth.

I. Network Components

There are four major components that make up a data processing network:

A. Data Base(s)
B. Application Module(s) Software
C. Hardware Component(s)
D. Supporting System(s) Software

A. Data Base(s)

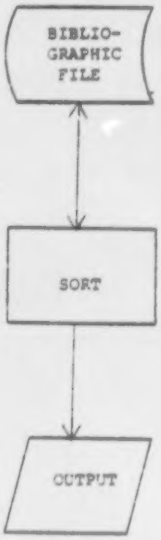
The data base is the most essential component of any system. It contains the various pieces of information needed in the processing of the application programs. A data base may contain:

1. A single file of like records (an example would be the OCLC bibliographic tapes);
2. Multiple files, none of which are interrelated;
3. Multiple files which are interrelated;
4. Multiple files and sub-files with some or all interrelated (See Figures 1a & 1b).

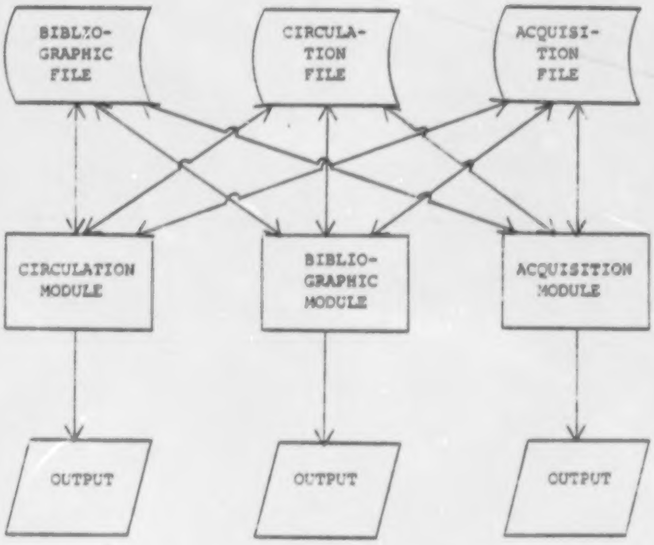
A fully integrated system would be of type 4 (Figure 1b); multiple files and sub-files with some completely

PROJECT/PROGRAM NAME FILE STRUCTURES	PROJECT NO	PAGE OF
CUSTOMER	PROGRAM NO	DATE

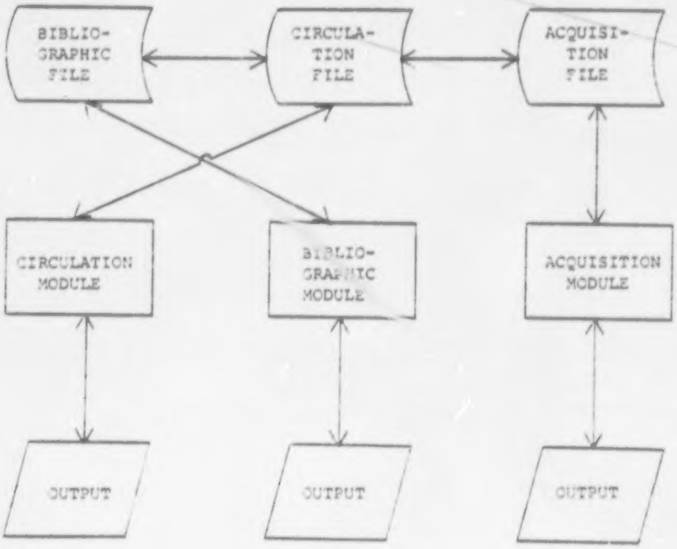
FLOW CHART


BIBLIO-
GRAPHIC
FILE
SORT
OUTPUT

SINGLE FILE-ONE PROCESS


BIBLIO-
GRAPHIC
FILE
CIRCULA-
TION
FILE
ACQUISI-
TION
FILE
CIRCULATION
MODULE
BIBLIO-
GRAPHIC
MODULE
ACQUISITION
MODULE
OUTPUT
OUTPUT
OUTPUT
OUTPUT

NON-RELATED FILES - MULTIPLE PROCESSES


BIBLIO-
GRAPHIC
FILE
CIRCULA-
TION
FILE
ACQUISI-
TION
FILE
CIRCULATION
MODULE
BIBLIO-
GRAPHIC
MODULE
ACQUISITION
MODULE
OUTPUT
OUTPUT
OUTPUT

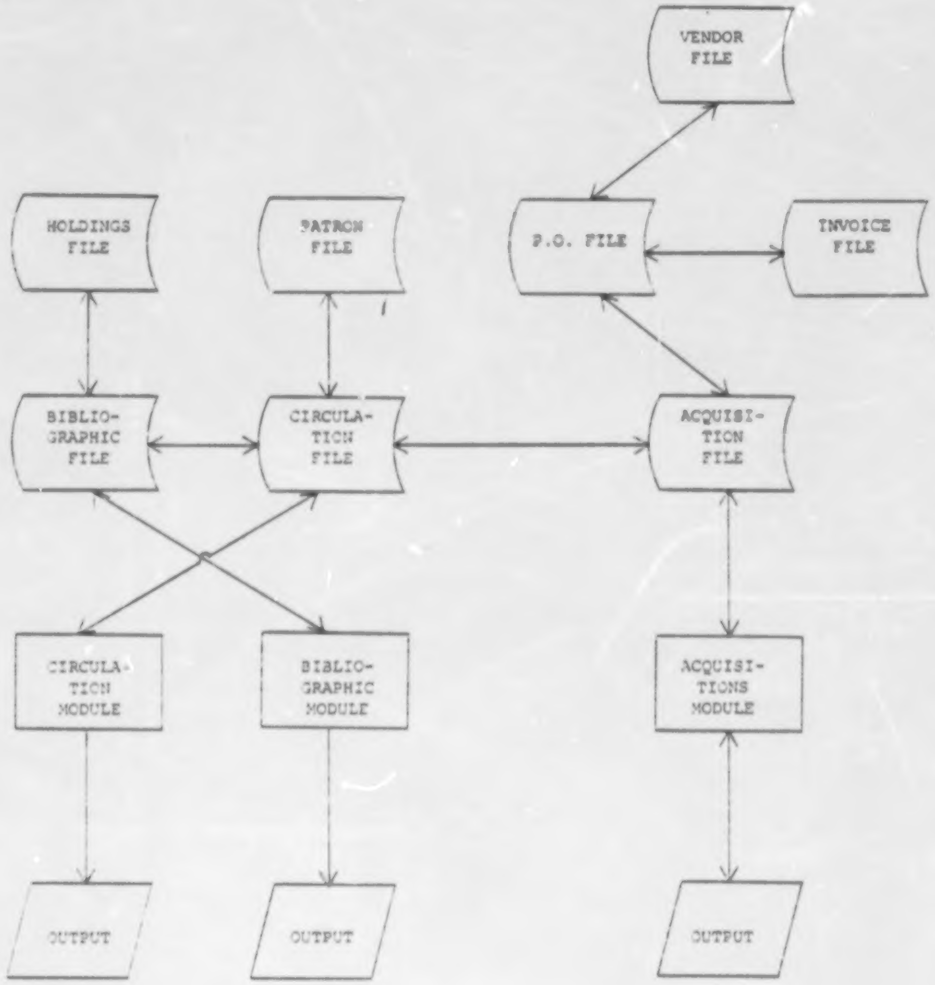
INTERRELATED FILES - MULTIPLE PROCESSES

FIGURE 1a

PREPARED BY ____________________

PROJECT/PROGRAM NAME	PROJECT NO.	PAGE OF
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART


VENDOR FILE
HOLDINGS FILE
PATRON FILE
P.O. FILE
INVOICE FILE
BIBLIO-GRAPHIC FILE
CIRCULA-TION FILE
ACQUISI-TION FILE
CIRCULA-TION MODULE
BIBLIO-GRAPHIC MODULE
ACQUISI-TIONS MODULE
OUTPUT
OUTPUT
OUTPUT

INTERRELATED AND NON-INTERRELATED FILES - MULTIPLE PROCESES

FIGURE 15

THE COMPUTER COMPANY - T961 (11/74)

PREPARED BY____________________

interrelated and some only specifically related. This type of structure is known as relational. Data is not repeated within the various files, but rather the data needed during the processing of any program module is obtained from the file containing this information, not necessarily from the file related to this module.

The contents, therefore, of an integrated system which would process:

1. Cataloging
2. Circulation - Reserves - Inter-library Loans
3. Acquisitions
4. Literature Searches

might be structured as shown in Figure 2.

This relational structure allows the system to process all the application modules, but more importantly:

1. Allows the user to obtain all the information needed through one query;
2. Allows all files to be updated from one transaction.

PROJECT/PROGRAM NAME RELATIVE FILE STRUCTURE		PROJECT NO.	PAGE OF
CUSTOMER		PROGRAM NO.	DATE

FLOW CHART

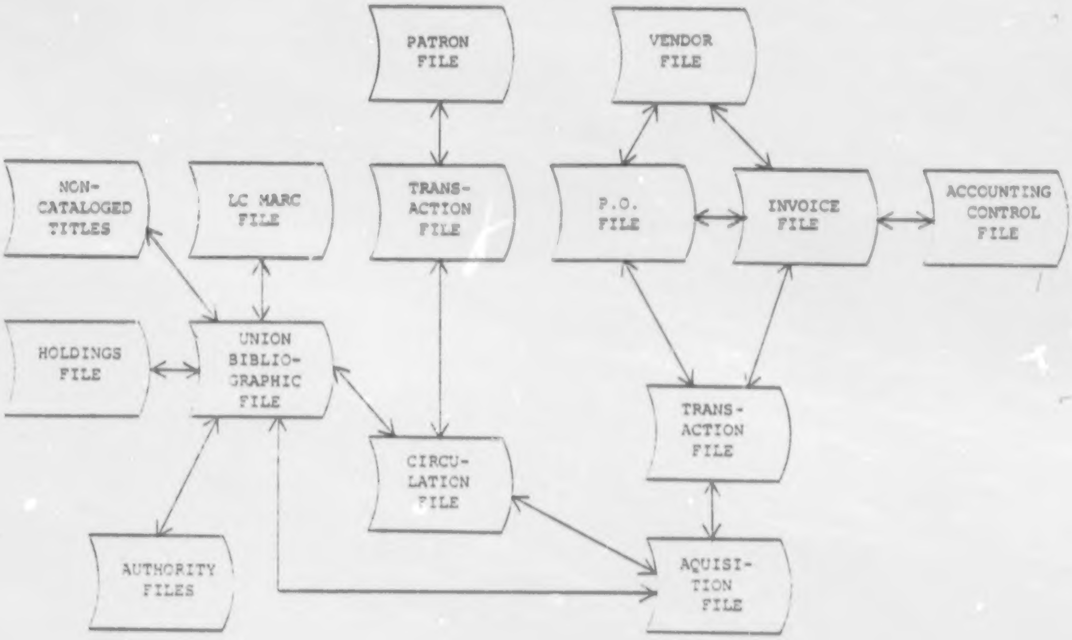

PATRON FILE
VENDOR FILE
NON-CATALOGED TITLES
LC MARC FILE
TRANS-ACTION FILE
P.O. FILE
INVOICE FILE
ACCOUNTING CONTROL FILE
HOLDINGS FILE
UNION BIBLIO-GRAPHIC FILE
TRANS-ACTION FILE
CIRCU-LATION FILE
AUTHORITY FILES
AQUISI-TION FILE

FIGURE 2

PREPARED BY

An example of this could be best illustrated through the use of the acquisitions module.

Before ordering a requested title the system could be queried about the title. (Figure 3)

1. Is it available in the region?
2. Is it already available in the library? If so, are there enough copies?
3. Is it already on order? If so, from whom and what is the status?

This one query should return all the information to the user.

It follows then that upon the receipt of an invoice, or even the initiation of a purchase order, the system should update all the files from the program involved. (Figure 4)

1. Acquisition Transaction File
2. Purchase Order File
3. Invoice File
4. Accounting Control File
5. Circulation File
6. Union catalog file (from LC MARC) or holdings file (from union catalog) or non-cataloged

PROJECT/PROGRAM NAME ACQUISITION QUERY	PROJECT NO.	PAGE OF
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

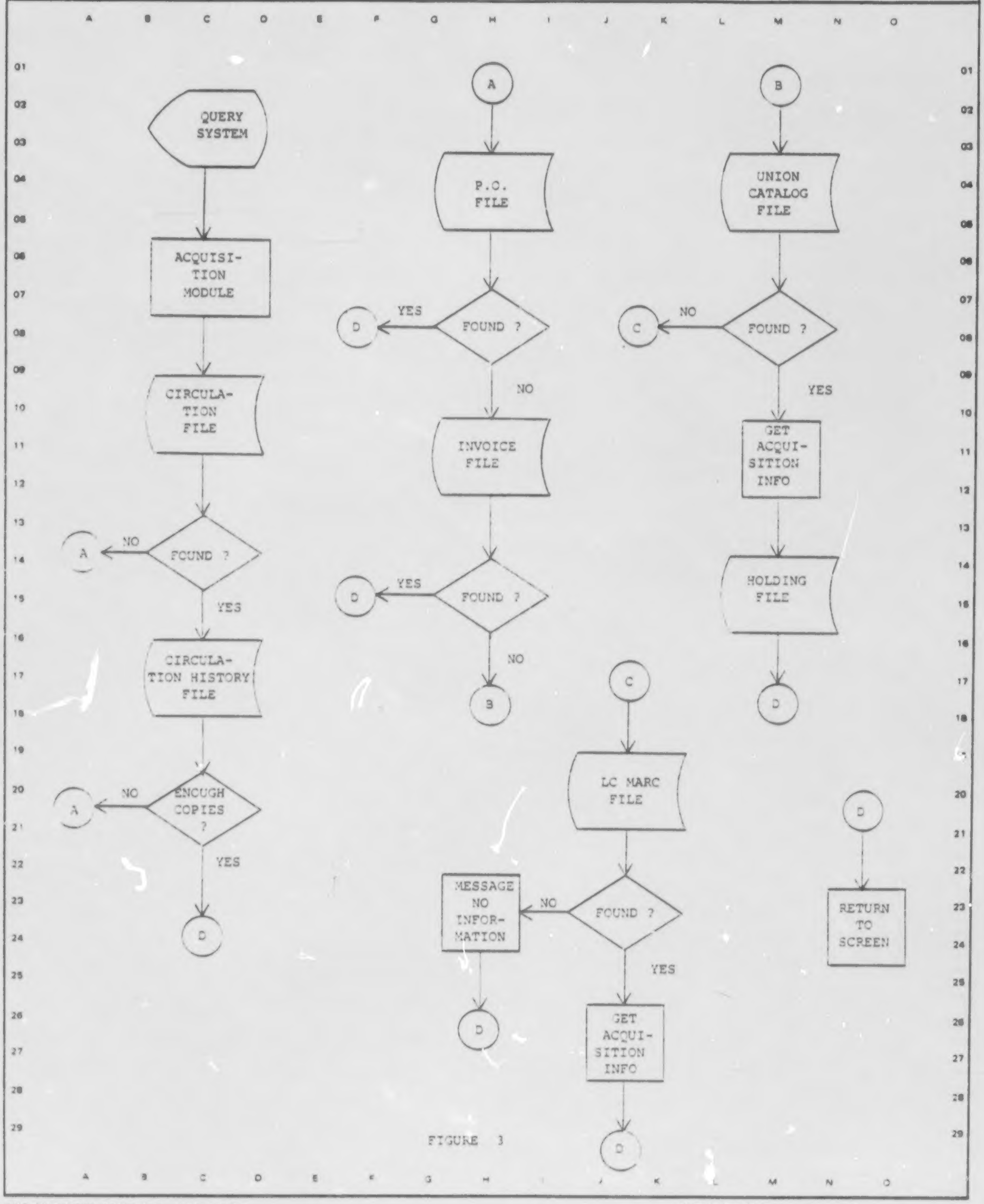

QUERY SYSTEM
ACQUISI-TION MODULE
CIRCULA-TION FILE
FOUND ?
NO
A
YES
CIRCULA-TION HISTORY FILE
ENOUGH COPIES ?
NO
A
YES
D
A
P.O. FILE
FOUND ?
YES
D
NO
INVOICE FILE
FOUND ?
YES
D
NO
B
B
UNION CATALOG FILE
FOUND ?
NO
C
YES
GET ACQUI-SITION INFO
HOLDING FILE
D
C
LC MARC FILE
FOUND ?
NO
MESSAGE NO INFOR-MATION
D
YES
GET ACQUI-SITION INFO
D
D
RETURN TO SCREEN

FIGURE 3

THE COMPUTER COMPANY - T961 (11/74)

PREPARED BY

PROJECT/PROGRAM NAME	PROJECT NO.	PAGE OF
ACQUISITION - UPDATE		
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

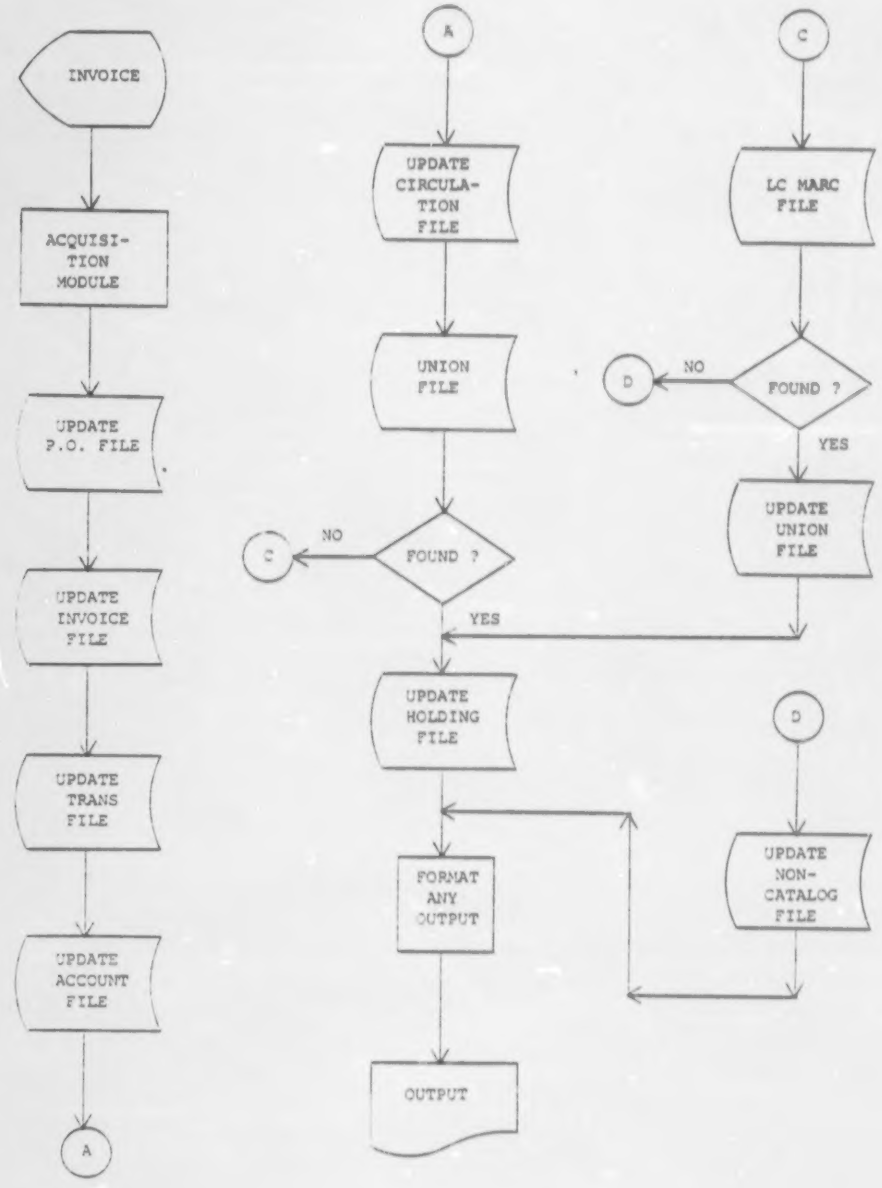

INVOICE
ACQUISITION MODULE
UPDATE P.O. FILE
UPDATE INVOICE FILE
UPDATE TRANS FILE
UPDATE ACCOUNT FILE
A
A
UPDATE CIRCULATION FILE
UNION FILE
FOUND ?
NO
C
YES
UPDATE HOLDING FILE
FORMAT ANY OUTPUT
OUTPUT
C
LC MARC FILE
FOUND ?
NO
D
YES
UPDATE UNION FILE
D
UPDATE NON-CATALOG FILE

FIGURE 4

THE COMPUTER COMPANY - T961 (11/74)

PREPARED BY

file (if title not in either the union or LC files).

Data Base Size: Once it has been decided which applications will be automated and, therefore, what files will make up the data base, an analysis must be made as to the expected size of each file, the expected growth rate of each file, the number of terminals to be used to access the data base, and the expected response time or turn-around time from these accesses. From this set of statistics, it is then possible to attempt to determine the hardware and software support required to efficiently process the data, both present and future. This type of analysis is also a major factor in determining the topological design of the network and the distribution of the data base.

To determine the size of the base, one must first determine the information to be contained in each file within the base. A file is a collection of like records; i.e., all records within the file contain the same type of information. For example, the typical MARC record can have only certain fields; however, the information within these fields pertain to an individual title (record) only.

Bibliographic File of MARC Records: The MARC format allows for variable length records and variable length fields within a record. The minimum length is 103 bytes (characters) and, by definition, there is no maximum. (Usually a maximum is set - OCLC has a current maximum of 5,000.) It has been found, however, that the average length is around 600 bytes. This figure of 600 multiplied by the number of unique titles would then give some idea of the amount of storage space needed for the file.

Statistics have been gathered within the State for the number of volumes for monographs only:

Public:	8,492,620
Academic:	10,163,537
Special:	4,591,847
School:	15,089,862
Total	38,337,866

To determine the number of unique titles within each of these collections, one must take the estimated ratio of titles to volumes for each segment:

Public: Ratio titles to volumes = .36

3,057,343 Titles

Academic: Ratio titles to volumes = .64

6,771,354 Titles

Special: Ratio titles to volumes = .81

3,718,395 Titles

School: There are no ratio statistics available. We shall assume the one which will give us the greatest yield: .81

12,222,787 Titles

This gives us a total of: 25,769,883 Titles

The problem now arises: of these 25,769,883 titles, how many are unique within the group? There are no statistics available to determine this. To attempt to do so manually would be a monumental job. For our purpose, we will assume that only 10% are unique which gives us a total of 2,576,988 unique titles which will represent the bibliographic file.

600 bytes x 2,576,988 = 1,546,192,800 bytes of storage for the file.

To obtain the growth rate factor, the following algorithm would apply:

Volumes x Growth Rate x Ratio Titles to Volumes = Segment Growth Projection

Total Segment Growth x .10 = Projected Total Growth Each Year

If we inject the following collected statistics into the algorithm:

SEGMENT	# VOLUMES	GROWTH RATE	TITLES/ VOLUMES
Public	8,492,620	.07	.36
Academic	10,163,537	.041	.64
School	15,089,862	.072	.81*
Special	4,591,847	.072*	.81

we obtain a growth rate per year figure for unique titles of 162,854 which translates to 97,712,460 bytes of storage. This figure, of course, would increase each year.

The same logic will be applied to each file within the data base regardless of the information contained in the file.

---

*No statistics available

Accessing Criteria: The actual accessing of the data within the base is determined by the applications to be processed. Certain applications would naturally fall into the on-line fast retrieval category (catalog searches, circulation, interlibrary loans). Others could be of a lower priority, with either a slower retrieval time or batch-processed during off hours (cataloging, acquisitions). The system should be able to accommodate all forms of data access and processing:

1. Batch (Off-line)
2. Batch (On-line)
3. On-line/Real-Time
4. Combinations of Above

and should be able to easily transfer from one mode to the other down to the individual program level, rather than simply to the application level.

Updating of Files: A truly integrated library data base will consist of extremely volatile files. This means that constant changes will be made to the records within each file. The data base is unique in that it is not static in size. Although one could liken it to an inventory, unlike most inventories, the product is not sold and, therefore, remains in the base. Even the client list (patrons) grows and changes.

The decision to update files in either a batch or on-line mode is determined in the same manner as was the access technique. It depends upon the timeliness of the file. The greatest concern relative to file updating should be in the structure of the file itself. All files within the base must contain:

1. Variable length records
2. Variable length fields within the record
3. Multi-keyed indexing

The operating and/or data base management system must be able to handle this kind of structure; allowing for change in record size, re-indexing of the file and re-assignment of available space in the case of any type of update (addition, change, deletion) to the file and it must be able to do so regardless of the mode of operation.

File Utilities - Merging, Purging, Adding, Conversion: The volatile, non-static nature of the library system file structure has been discussed (File Accessing). It was determined that in order to handle such a data base, the internal file structure of the base must contain:

1. Variable length records
2. Variable length fields within the record
3. Multi-keyed indexing

These three criteria must also be addressed when dealing with file utilities.

Conversion/Merging: There are libraries in the State which have data (cataloging, circulation, acquisitions) in some type of machine readable form. To manually re-enter this data would be not only time consuming, but would not be cost-effective. This procedure can be performed more easily and in a more cost effective manner by use of the system itself. However, certain criteria must be set out for such file conversion:

1. Program:
    a. Must be able to convert from one file structure to another
    b. Must keep good controls of input and output
2. Operating and/or data management system must be able to efficiently merge the incoming records into the existing files - handling the variable length record, field, and re-indexing of the file.

Purging: Purging of files, whether in total or in part, must also conform to the variable length and multi-keyed concept. This, again, for efficiency must be handled by the operating and/or data management system and not by the program itself.

Adding: A fully integrated system cannot be established overnight. To be effective there must be an overall design concept. However, it is impossible, even under ideal conditions, to anticipate all the new applications or system modifications which may be desired subsequent to the original design. It is, therefore, imperative that the design allow for easily adding new modules and associated files as the system expands. As was stated in the beginning of the section, it is essential that the data base be relational and not relative, therefore, each added file will in some way be relative to already existing files. It cannot be stressed too strongly that this concept, along with the unique structure criteria of the data base, represents the greatest problem in system design and development. The adding of data files which in turn lead to restructuring of the data base must be efficient. There are very few operating and/or data management systems that will handle this without a great deal of work on the part of the people in charge of the network system. Any operating and/or data management system considered must be able to handle the addition of a file and the linkage between it and any other related file(s) easily and efficiently. (See section D of main study.)

Security and Data Control:

Internal Control

Figure 2 represents the file structure of the network in a form such that all like data, regardless of the source, resides in the same file. In this structure the problem of internal security control is less acute than if the structure consisted of multiple smaller files, one for each user, with nonshared unique information (Figure 5). Notice that User 1 and User 2 are sharing the bibliographic module, but each has opted to maintain separate circulation files (with the exception of the patron file) and acquisition files (with the exception of the vendor file). Ignoring the dashed lines, one can see that even in this configuration the files are completely relational. Suppose, however, that User 1 and User 2 do not wish to share their circulation transactions file nor their internal accounting file. To provide this type of security within an application program would require that each time either user wished a secured file the program would have to be changed, and to initiate it at the terminal level is not possible since most files are shared. Again, this type of control must be provided by the operating and/or data management system and it must be provided with no down-time due to reorganization of the files.

Internal data control capability must also be provided by the operating and/or data management system. It must be able to detect and eliminate data redundancy, detect and report any file structuring problems, file overload problems, and data content problems. The user must be made aware of these

PROJECT/PROGRAM NAME MULTIPLE FILES	PROJECT NO.	PAGE OF
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

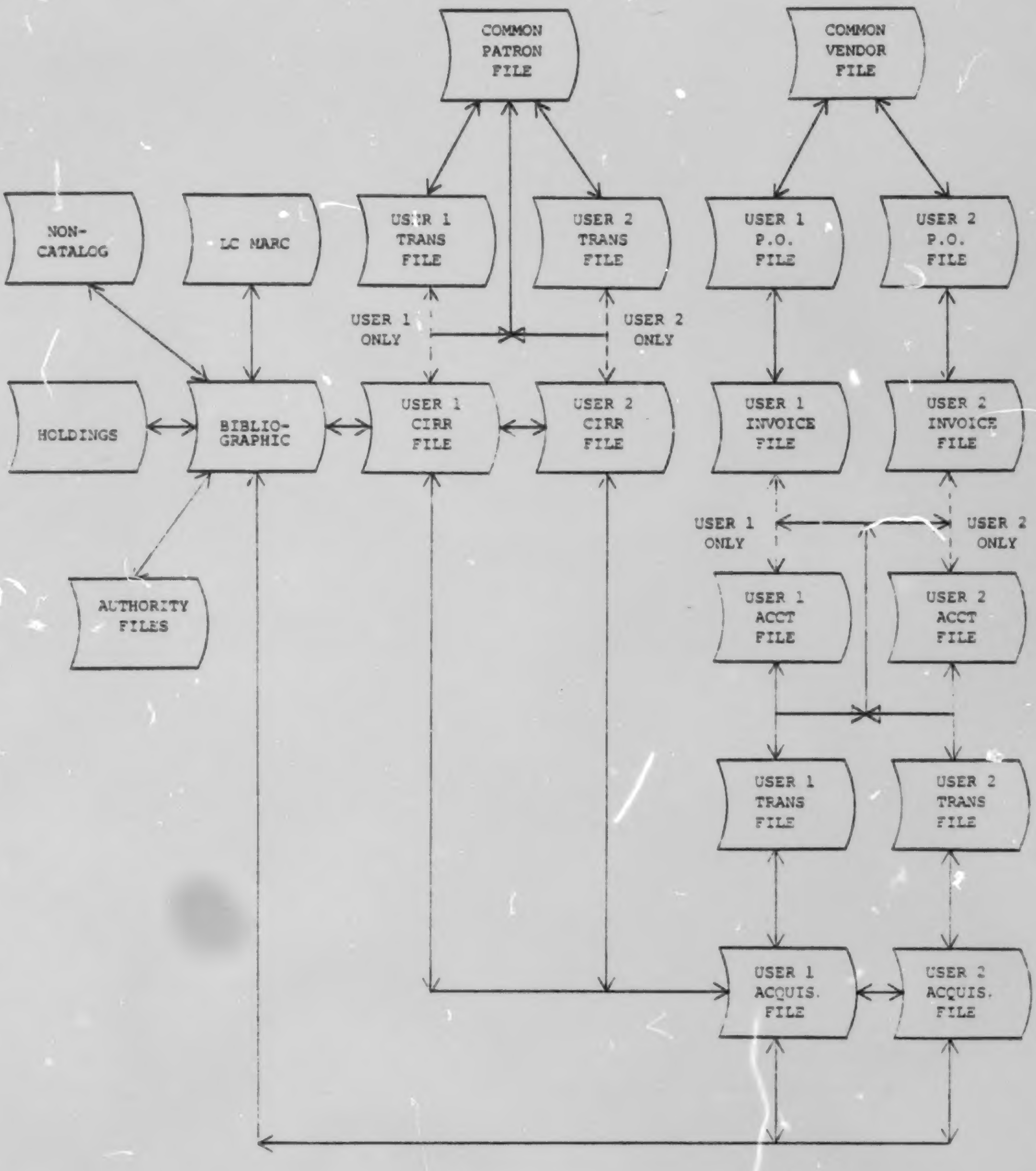

COMMON PATRON FILE
COMMON VENDOR FILE
NON-CATALOG
LC MARC
USER 1 TRANS FILE
USER 2 TRANS FILE
USER 1 P.O. FILE
USER 2 P.O. FILE
USER 1 ONLY
USER 2 ONLY
HOLDINGS
BIBLIO-GRAPHIC
USER 1 CIRR FILE
USER 2 CIRR FILE
USER 1 INVOICE FILE
USER 2 INVOICE FILE
USER 1 ONLY
USER 2 ONLY
AUTHORITY FILES
USER 1 ACCT FILE
USER 2 ACCT FILE
USER 1 TRANS FILE
USER 2 TRANS FILE
USER 1 ACQUIS. FILE
USER 2 ACQUIS. FILE

FIGURE 5

THE COMPUTER COMPANY – T961 (11/74)

PREPARED BY

problems (the means would depend upon the mode of operation) before any modification is made to the data base by the system itself or any controlling program.

External Control

External security controls must also be present in the design. These controls must be easily changed by system management without requiring program changes. Passcode security into the system must be multi-leveled and changeable by individuals designated by each user.

External data checking should be done in a conversational mode; i.e., error message given to the user at the time the erroneous information is keyed - not after an entire screen of data has been transmitted. (This type of interaction could only occur in an on-line mode of operation).

Processing of the Data Base: To design a fully integrated library network, two basic premises must be kept in mind:

1. It must satisfy the needs of a variety of users and do so in the most cost-effective way for each individual user. There will be those who will need and will take advantage of all the power the system can provide; but conversly, there will be those who do not need all of the system's capabilities.
2. The network will be a growing entity. New users will be added and some users may decide not to continue as part of the network. The population within user groups will be changing and, along with this, the needs that the user library must provide to the population.

With this in mind, the data base itself must be thought of as volatile and non-static. Not only must it be available to the user through a mix of operational modes (batch, on-line, real-time) but the data structure must be such that it can be easily distributed wherever and whenever the need arises. This distribution could be to other processors or could be a redistribution within a central or satellite processor. (See Figures 6 & 7)

From this narrative on the data base it can be concluded that the following features must be present or planned for at the conceptual design stage:

1. The number, content and approximate size of each file and sub-file that will eventually make up the data base must be determined.

PROJECT/PROGRAM NAME CENTRALIZED PROCESSING	PROJECT NO.	PAGE OF
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

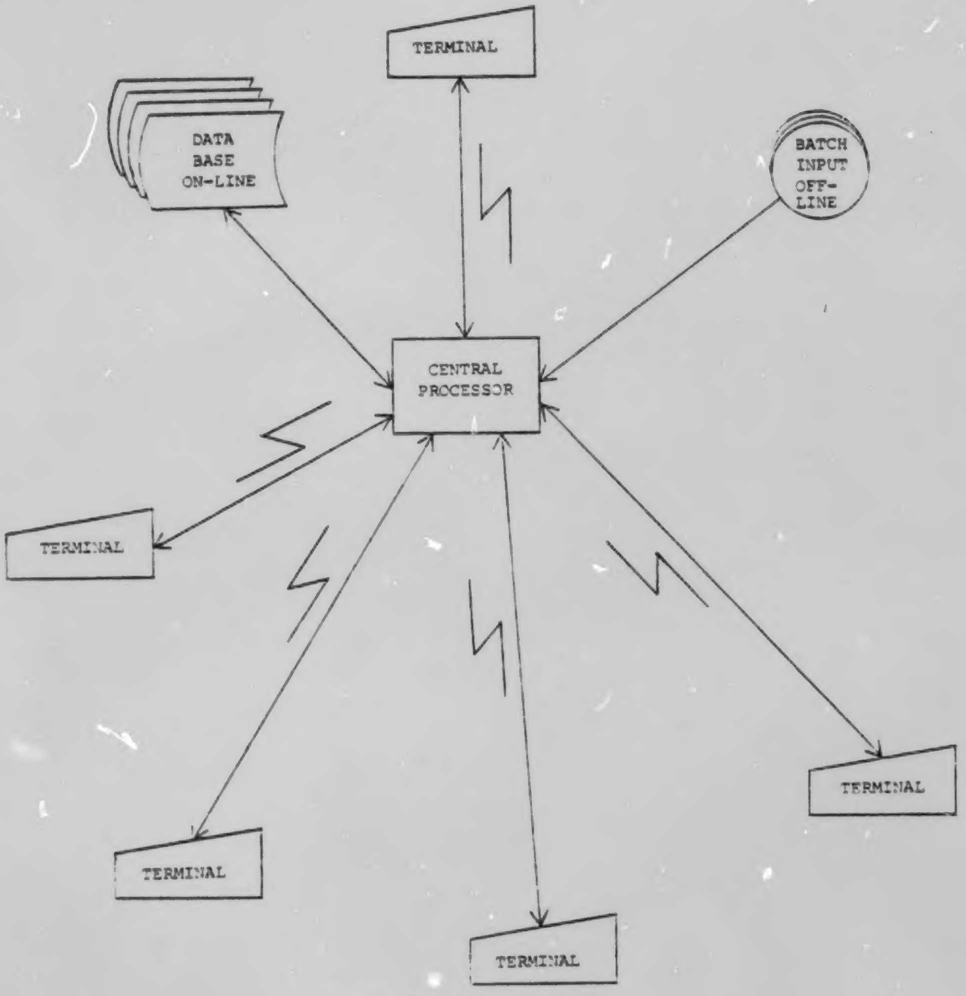

TERMINAL
DATA BASE ON-LINE
BATCH INPUT OFF-LINE
CENTRAL PROCESSOR
TERMINAL
TERMINAL
TERMINAL
TERMINAL
TERMINAL

FIGURE 6

THE COMPUTER COMPANY – T961 (11/74)

PREPARED BY ____________

PROJECT/PROGRAM NAME DISTRIBUTED PROCESSING	PROJECT NO.	PAGE OF
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

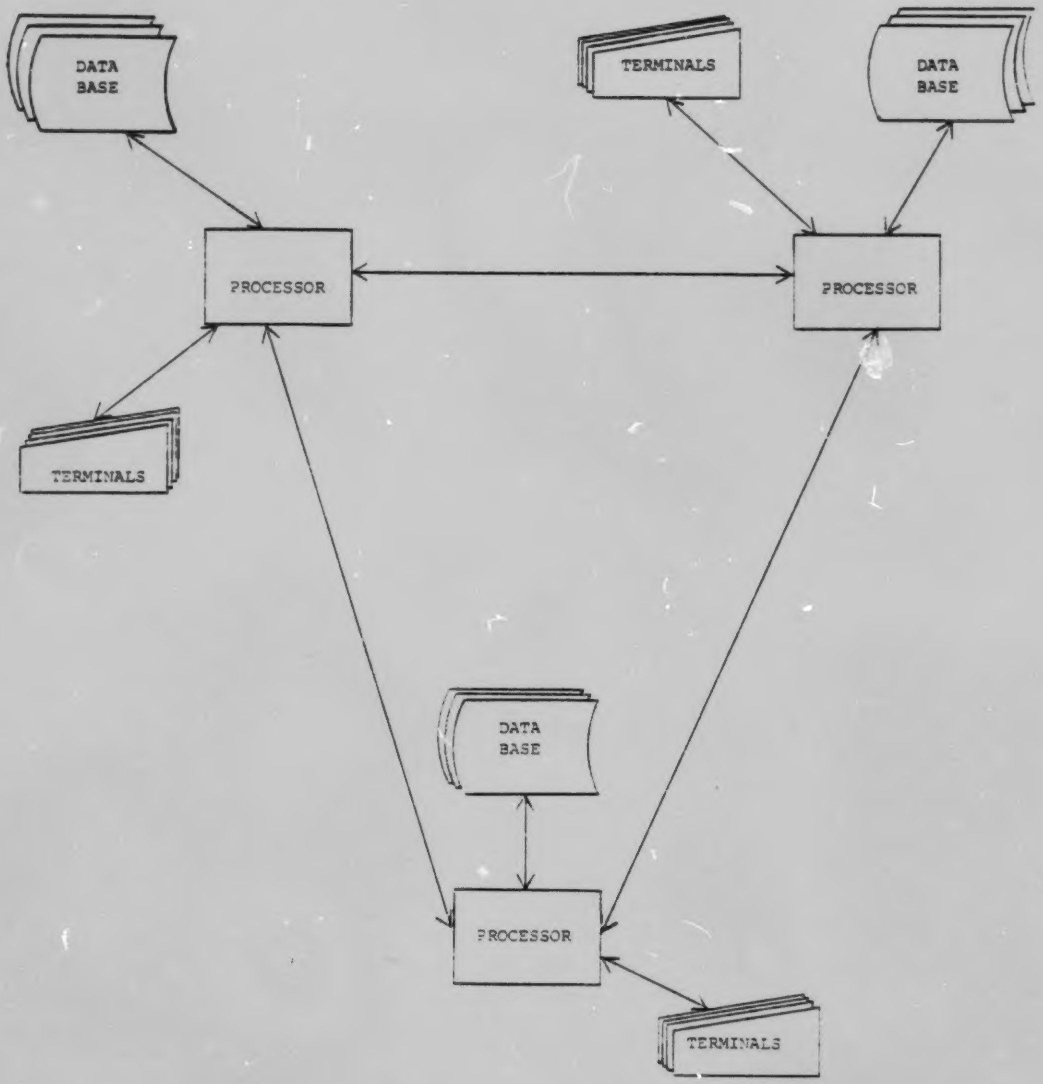

DATA BASE
TERMINALS
DATA BASE
PROCESSOR
PROCESSOR
TERMINALS
DATA BASE
PROCESSOR
TERMINALS

FIGURE 7

THE COMPUTER COMPANY – T961 (11/74)

PREPARED BY ____________________

2. The data base must have a relational rather than a relative structure.

3. The operating and/or data management systems must be able to handle:

    a. Variable length records

    b. Variable length fields within a record

    c. Multi-keyed indexing of the file

    and be able to do this regardless of the mode of operation:

    a. Batch (Off-line)

    b. Batch (On-line)

    c. On-line/Real-Time

    d. Combination of any of the above

4. The operating and/or data management system must be able to perform all data base maintenance, allowing for change in record size, re-indexing of the file, re-assignment of available space, and changes in linkages between the files.

5. Mechanisms must be available for the easy use of file utilities such as:

    a. File merges

    b. File purges

    c. Addition of new files

    d. Conversion of files to the system data structure

6. Internal and external data control must be present and easily applied.

7. The volatile and non-static nature of the data base must be kept upmost in mind during all planning, design and implementation of the integrated system.

B. Application Modules

At this point in the planning of an integrated library system, it is uncertain from which base the network will develop. It could develop from a centralized holding of bibliographic records, with each regional area using its circulation records (upgrated to full MARC) as input to update the centralized holding. It could also develop from separate regions using turn-key and vendor offered applications. Section C of the study deals in depth with the recommended approach to the network. This addendum will deal with the data processing approach to each application module, addressing the following needs for each:

1. Data Content
2. Input Methods
3. Processing Methods
4. System Integration Requirements
5. Available Systems
6. Time and Cost Projections
7. Topological Configurations

From this analysis of the various application modules, we will draw the conclusions and recommendations for the network approach and the administration of such a network.

Cataloging

The basis for all applications indigenous to a library is the base of bibliographic records. This does not mean that the network must begin here but, regardless of which applications are addressed first, they must be able to supply input to upgrade to a full bibliographic MARC format for creation of the bibliographic base. (See Section 3&A Study).

From the findings in the study, we know that some libraries are using commercial services, such as OCLC for bibliographic searches and production of catalog cards. Others have their holdings in some machine readable form which could be used as input to a bibliographic file such as offered by Brodart, Blackwell, etc., to obtain full bibliographic records.

The State Library has contracted with a local vendor to produce a union catalog of holdings using the OCLC archival tapes from 22 academic libraries. The processing of these tapes, some of the problems which occurred, how they were solved, and problems which must still be resolved can provide insight in the design of the library system.

It was decided that two types of catalogs would be produced from the tapes:

1. Union Catalog - containing the records in alpha sequence by:
    a. Main entry
    b. Author entry
    c. Added entry
    d. Subject entry
2. Shelf listing sort by classification code for each individual library using main entry only.

Due to the difference in cataloging format for the various entry types, five different formats were decided upon:

1. Main entry - union
2. Main entry - holding
3. Author entry
4. Added entry
5. Subject entry

The first processing run answered the question regarding duplicate records. It was found that for each library holding a given title, there was a record present in the data base. It was also discovered that there were duplicate records even within a given library.

In order to obtain an estimate of processing time, requirements and ratio of input records to output records, a sample of 2500 records was processed on an IBM 370/145. The following statistics were gathered:

1. To format the key, it took 1 minute CPU time
2. To sort the file took 1 minute CPU time
3. To reformat each input record into the various output formats, took 20 minutes CPU time
4. The ratio of input record to output record was 1 input - 3.5 output.

The timing statistics do not seem unreasonable until one realizes that we are discussing CPU time and not clock time. It would certainly not be cost-effective to use such a large machine for this process alone, so it must compete with other processes at the same time. The total CPU time of 22 minutes translates to 1.1 hours of clock time.

If we assume there are 250,000 input records, then the processing time becomes quite significant:

22 min. x 100 = 2,200 min. or 36.6 hours CPU

1.1 hrs. x 100 = 110' or 4.6 days clock time

Input records = 250,000 - Output = 875,000

From these statistics, it was quickly determined that in order to be cost-effective, two significant changes would be necessary in the program design and system processing techniques.

1. Some portion of the program must be written in machine level code rather than COBOL.
2. Some type of data base handling software is needed in order to greatly speed up the processing of the output record formats. (See section 3 & D of Study)

It was also discovered that the records are in various natural languages and that OCLC has special characters and unique combinations of characters on its terminal in order to accommodate them. Therefore, the COM processor must be equipped with an ALA print character set and the special characters programmed into the COM processor in order to appear on the output microform. Some other areas of concern, which are not actual programming problems, but which must be determined by the State before further processing can take place are:

1. What to do about duplicate records. If several libraries hold the same title, which entry should be maintained in the base? If these duplicates are deleted and a sub-file created containing information needed for each individual library's shelf listing, what information should be kept in this file pertaining to the individual cataloging by the library?
2. Each individual library has a different hierarchy in assigning classification numbers. These must be obtained before shelf listings can be produced.

Data Content

At least two files are necessary to form the data base, regardless of the mode of operation (figure 2):

1. Bibliographic file
2. Holdings file

Strong arguments can also be presented for the existence of authority control files. (See Page 40 - On-line Cataloging.) Before these can be integrated into the base, however, programs must be written to search and report from the base the number of discrepancies existing within the file. If accomplished, however, all succeeding entries into the base could be processed against the authority files regardless of mode of operation.

If the ratio of 10% is correct to determine unique titles, then 10% of 250,000 (OCLC records) is 25,000. This is certainly a small portion of the projected unique titles within the State (2,576,988). Section 3 of the study proposes that the data base be enhanced by converting the available machine readable circulation records and microform catalog tapes into full MARC records and along with this, begin the retrospective conversion of existing card catalogs to tape form and finally to full MARC. There are several commercial firms which offer this service.

Before embarking on such a course of action, which at best is costly in time and money, certain questions must be thoroughly investigated and answered by the State:

1. What common information, which can be used to match against a bibliographic base, is present in the existing catalog to be converted?
    a. LC Card Number
    b. ISBN or ISSN Number
    c. Author/Title (full or shortened version)
2. Is there personnel available to undertake such a conversion?
    a. Data entry clerks

	PROJECT/PROGRAM NAME BATCH - CATALOG PROCESSING	PROJECT NO.	PAGE OF
	CUSTOMER	PROGRAM NO	DATE

FLOW CHART

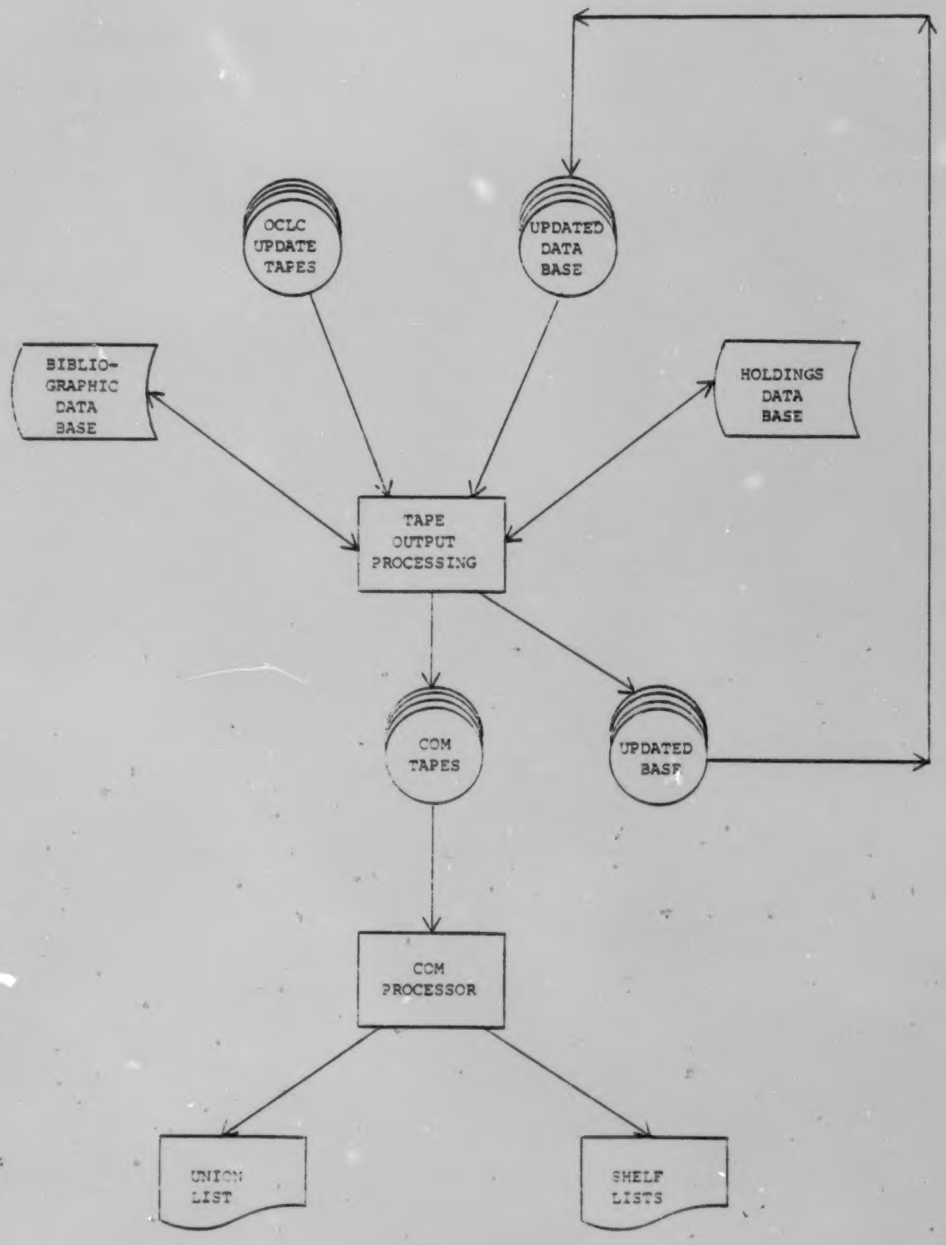

OCLC UPDATE TAPES
UPDATED DATA BASE
BIBLIO- GRAPHIC DATA BASE
HOLDINGS DATA BASE
TAPE OUTPUT PROCESSING
COM TAPES
UPDATED BASE
COM PROCESSOR
UNION LIST
SHELF LISTS

FIGURE 8

THE COMPUTER COMPANY – T961 (11/74)

PREPARED BY

   b. Professionals to check retrieved cataloging against the original

3. What will be the mode of data gathering?
   a. Typing on special paper for optical scanning
   b. Direct input to magnetic tape
4. What has been the experience of 'hits' by those libraries already involved in such a venture?
   a. Are their holdings similar to the library being converted?
   b. What bibliographic data base(s) does the vendor use for matching?
   c. What is the quality of the cataloging?
5. What is the cost of such a conversion?
   a. Are there funds available?
   b. Will the vendor give exact costs?
6. Would it be more cost-effective for a group to combine and create its own programs for conversion?
   a. What would be the mode of input?
   b. Number of personnel needed
   c. Programming and hardware costs
   d. Possible use of State data base for retrieval of bibliographic records
7. How will the cataloging be done for those titles not found?

Figure 9 sets out in schematic form the normal process of conversion used by commercial vendors.

PROJECT/PROGRAM NAME
RETROSPECTIVE BATCH CONVERSION
CUSTOMER
PROJECT NO.
PROGRAM NO.
PAGE OF
DATE

FLOW CHART

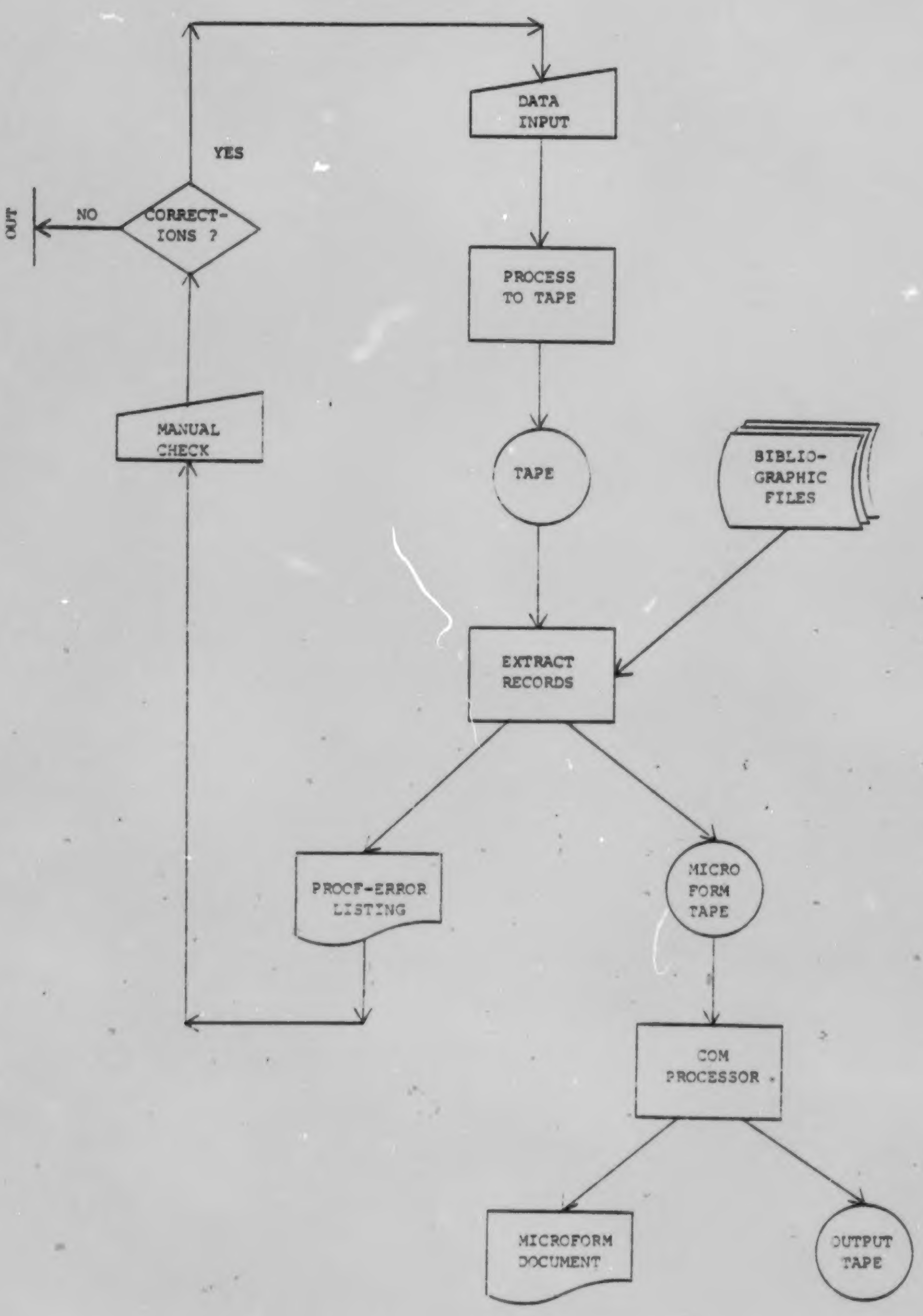

DATA INPUT
YES
OUT
NO
CORRECT-IONS ?
PROCESS TO TAPE
MANUAL CHECK
TAPE
BIBLIO-GRAPHIC FILES
EXTRACT RECORDS
PROOF-ERROR LISTING
MICRO FORM TAPE
COM PROCESSOR
MICROFORM DOCUMENT
OUTPUT TAPE

FIGURE 9

THE COMPUTER COMPANY – T961 (11/74)

PREPARED BY:

This process must be performed completely in batch mode. The user has the option to decide what information to use as input for comparison with the vendor's data base of bibliographic MARC records. This information will normally consist of:

1. LC card number
2. ISBN number
3. ISSN number
4. Author/Title (short form)
5. Author/Title (long form)
6. Combination of any of these

The vendor provides the format for the user and the information is typed onto special paper using a Selectric typewriter and a special print ball. The pages are sent to the vendor who transfers them to magnetic tape by use of an optical character scan mechanism. These tapes are then read into the vendor's main processor and the bibliographic files are searched for each entry. When a hit is made, it is written to tape. A print-out of the tape is returned to the user (this print-out may be on different media depending upon the vendor: paper, catalog card, etc.) for manual checking. Some statistical reporting is also produced to inform the user of the entries which did not result in the retrieval of a MARC record. The user informs the vendor of any errors; the vendor deletes these from the user's master tape and usually returns a microform catalog of the holdings along with the tape of bibliographic records.

As new holdings are acquired, the same process will be used and the records retrieved will be merged with the original tape.

Figure 10 is a schematic of an on-line conversion and cataloging system that could be developed by the State. Its use can be two fold:

1. Retrospective conversion
2. On-line cataloging

Input to this system would initially be through the use of portable on-line cathode ray tubes (CRT) using dial-up telephone lines for communication to the central processor.

The program controlling the input should have the capability of performing edit and validation checks and also should allow the user to correct input on-line. The information entered should be routed

PROJECT/PROGRAM NAME ON-LINE CONVERSION		PROJECT NO.	PAGE OF
CUSTOMER		PROGRAM NO.	DATE

FLOW CHART

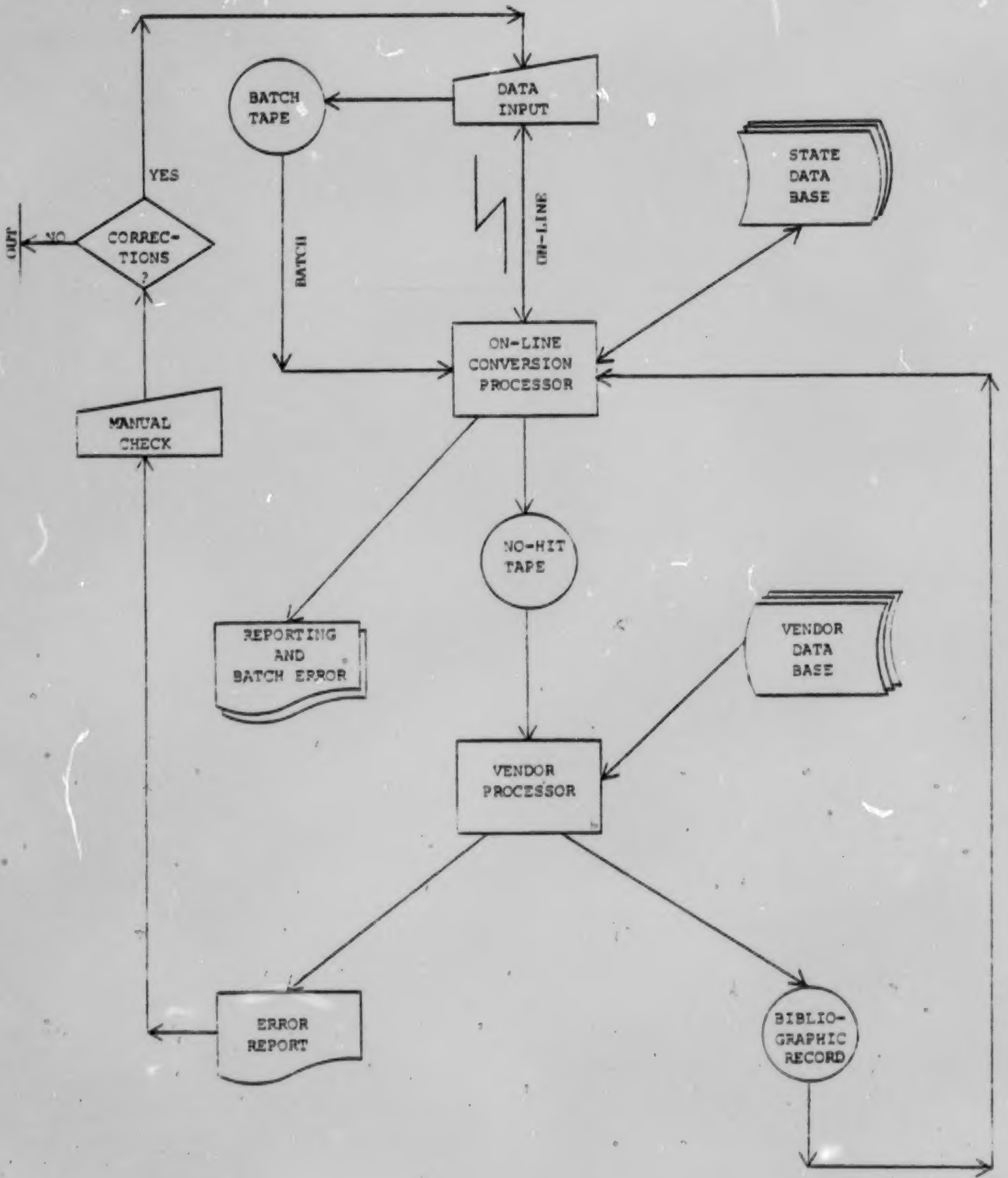

DATA INPUT
BATCH TAPE
STATE DATA BASE
YES
NO
OUT
CORRECTIONS ?
BATCH
ON-LINE
ON-LINE CONVERSION PROCESSOR
MANUAL CHECK
NO-HIT TAPE
REPORTING AND BATCH ERROR
VENDOR DATA BASE
VENDOR PROCESSOR
ERROR REPORT
BIBLIOGRAPHIC RECORD

FIGURE 10

THE COMPUTER COMPANY – T061 (11/74)

PREPARED BY ____________________

directly to magnetic tape for processing against an outside vendor's bibliographic data base. The program should be able to search the State data base and write information to tape only if the matching record is not found in the on-line data base. If found, the record should be displayed to the user to input holding information to be added directly to the holding file. The programs involved in the processing should keep the proper statistics needed for meaningful reporting to the user library and to the entire network.

Processing in this manner would be:

1. Faster, more efficient, and produce fewer errors
2. Make good use of the existing data base as a retrospective cataloging tool in an on-line environment
3. Provide cost savings, especially as the data base grows, by eliminating payment to an outside service for the duplication of cataloging of titles already in the State base.
4. This same process should be useable for adding holdings information of newly acquired titles by user libraries; and act as an aid in the acquisition of titles.

The logic involved in the programs contributing to this process will be useful during the entire history of the network and can provide the basis for the on-line cataloging module.

The process must be able to run in batch (off-line), batch (on-line) or on-line/real-time modes.

By adding LC MARC data to the State bibliographic data base, additional cost savings can be realized.

When this process is completed, the network will have a data base consisting of:

1. Bibliographic data derived from:
    a. OCLC tapes
    b. On-line and batch conversion of holdings
    c. Records retrieved from the use of machine readable tapes produced as a by-product of circulation, acquisition, microform catalog systems, etc.
2. Authority files
3. LC MARC files

Figure 11 is a schematic of the network at this point. Notice that updates to the base are still being received from bibliographic vendors, such as OCLC or Ballots. This interaction would probably exist for some time (See Section 3 of study).

PROJECT/PROGRAM NAME	PROJECT NO.	PAGE OF
NETWORK SCHEMATIC - FIRST PHASE		
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

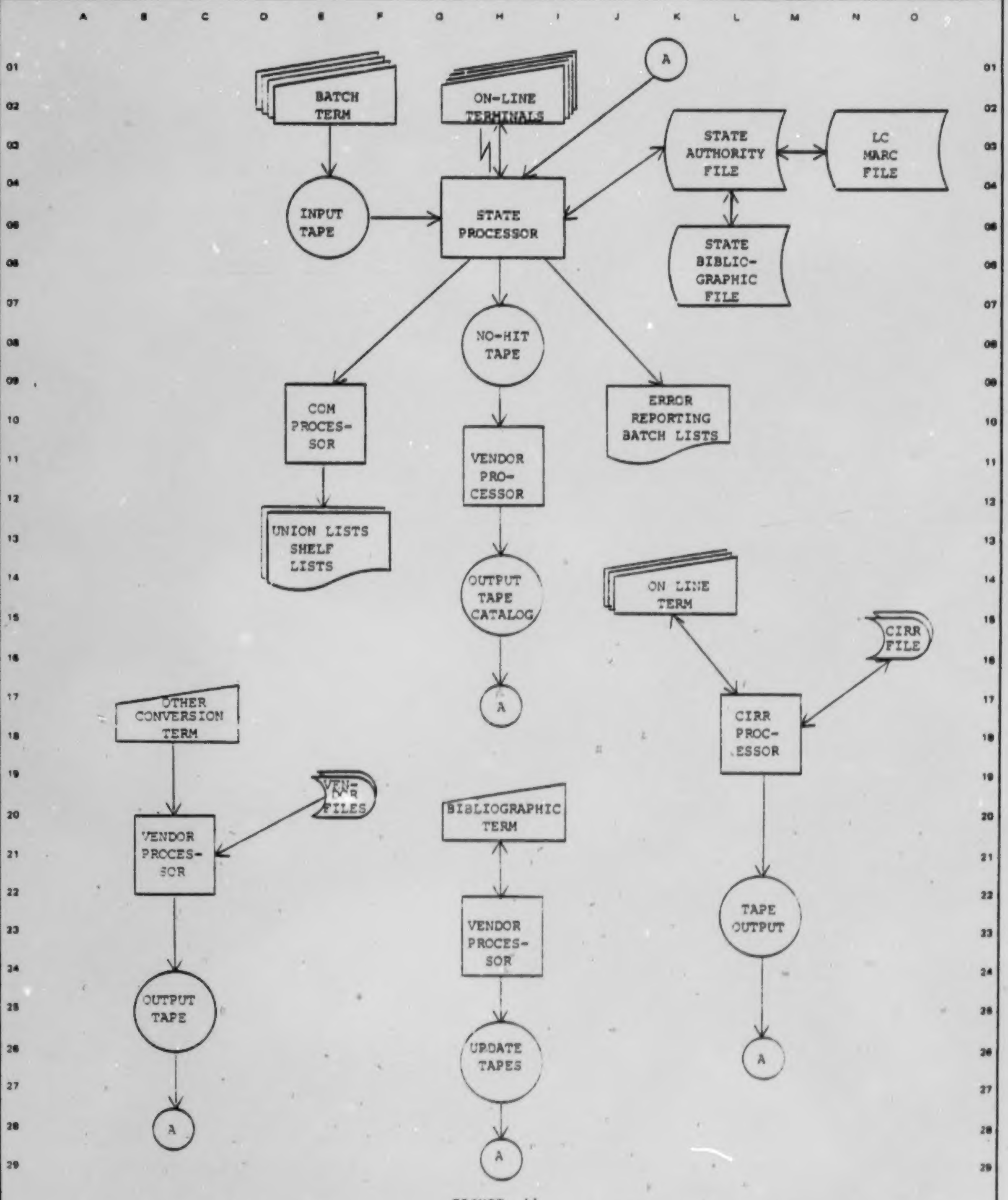

A
BATCH TERM
ON-LINE TERMINALS
STATE AUTHORITY FILE
LC MARC FILE
INPUT TAPE
STATE PROCESSOR
STATE BIBLIOGRAPHIC FILE
NO-HIT TAPE
COM PROCES-SOR
ERROR REPORTING BATCH LISTS
VENDOR PRO-CESSOR
UNION LISTS SHELF LISTS
OUTPUT TAPE CATALOG
ON LINE TERM
CIRR FILE
A
OTHER CONVERSION TERM
CIRR PROC-ESSOR
VEN-DOR FILES
BIBLIOGRAPHIC TERM
VENDOR PROCES-SOR
VENDOR PROCES-SOR
TAPE OUTPUT
OUTPUT TAPE
UPDATE TAPES
A
A
A

FIGURE 11

PREPARED BY ____________________

Processing

The next step in the evolution of the cataloging module would be to allow for on-line cataloging into the bibliographic base. If all criteria discussed in the section on Data Bases have been adhered to, the operating and/or data base management system chosen has the capability to handle:

1. Variable length records
2. Variable length fields within a record
3. Multi-keyed indexing

and the data base structure is relational so that each file within the base may be accessed from any other file.

Program(s) Requirements: File Maintenance

On page 16 of this addendum there is a discussion of file maintenance in relation to data structures. Here, as with all succeeding application modules, the discussion will be a programmatic approach to file maintenance. Programmatic file maintenance routines are:

1. Addition of a record to a file
2. Changing of a record in a file
3. Deletion of a record from a file

Addition of a Record to a File

The program should perform the following functions:

1. Determine from input criteria if the record already exists. (The requisites for this would be determined during the design stage.) If it does - display back the retrieved record(s) to the user for verification (some override must be accepted).
2. Interface with whatever mechanism is being used by the system to edit and check field content. (If field must be numeric, etc.) and display back to the user any errors made in input.
3. Allow for changing of any data before record is filed.
4. Interface with any authority files present in the system for subject entry, serial control, etc., and allow user override if necessary.
5. Interface with all internal and external security controls within the system.
6. Perform all validity checks of interacting data.
7. Format the record into the correct data structure for the file.
8. Interface with the data base management system for the restructuring of the file once the record is ready to be added.

Changing of a Record in a File

1. Determine from input if the record does exist. If it does not, display the proper error message. If the record(s) is found, display each in turn to user for validation.

   (3)-(6) Would be the same as the addition function.

   (7) Reformat the record into the correct data structure for the file.

   (8) The same as addition function.

Deletion of Record from a File

1. Determine if record can be deleted from the file. (Its relationship to other files). If it can be deleted - display the retrieved record for user final action.
2. Interface with data management system for restructuring of the file once deletion has taken place.
3. Interface with all internal and external security controls within the system.

Transaction Processing

In the processing of the bibliographic data files, the created transactions should provide for two functions within the system:

1. System Back Up - The transaction record must carry enough information to be used to recreate the action performed on the data file (addition, change, deletion) in case of a system failure. The transaction should be written to a separate storage

area (disk or tape) in a real-time mode. These files may subsequently be used to update the data files without the need for rekeying.

2. System Control and On-Line Analysis - Not as essential as system back-up, but still very advantageous for overall system control and analysis, is the updating from the transaction of control files pertaining to the cataloging function. The data collected for later analysis might include such statistics as:

    a. Number and type of additions to the file

    b. Number of changes

    c. Number of deletions

    d. Any other statistics that might be determined as useful in obtaining quick information about the bibliographic data base.

Form and Content of Output

The output format programs of the cataloging module should allow for:

1. Display via the CRT
2. Printing on a catalog card
3. Microform Output
4. Control record display and printing
5. Statistical reporting

The form(s) of output by each individual user should be set by a changeable, parameter-driven record; i.e. depending upon the values given, the system will or will not produce cataloging cards, microform version of the catalog (in whatever format(s) user designates), etc. These parameters should be changed as needed by the user and/or system supervisor. The content of the output should also be parameter-driven and changeable by the user and/or system supervisor; i.e., full MARC record display/print; certain fields from the record display/print, etc., or any combination of output formats.

Hardware Requirements

The hardware requirements of the cataloging module will depend upon the extent of its use. For instance, if its use is limited to a batch environment and only for the purpose of producing, in microform, a union listing of holdings for those libraries using the services of OCLC, the supporting hardware would be quite different in configuration than if the system is used in an on-line mode and integrated into the other functions that might be available through the use of the library system.

If on-line cataloging along with the many functions associated with it is to become a reality, the hardware must be easily upgraded and expanded to allow for growth. The LC MARC file alone consists of nearly one million records for monographs only. The serials file is close to 300,000. Virginia has a rather extensive group of holdings, especially in the academic area, that would not be included in these files and would be resident in the common holdings file.

The hardware must be dedicated to the processing of the library system and must be administered by an autonomous group who in turn are answerable only to the governing body of the network.

Software Support Requirements:

Data Base Management System

A decision regarding the type of file management or data base management system to be used is dependent upon the use being made within the system of the catalog (bibliographic) data file(s).

Batch (Off-Line) Processing - Single File

It was discovered by the vendor producing the union list of holdings from the OCLC archival tapes that in order to obtain any amount of efficiency some type of file management software (not necessarily data management software - See Section 3 of study) must be incorporated into the processing of the catalog. In this type of batch (off-line) single file processing, the program alone controls the interaction between the file and any supporting file management software. The determinate factors as to the restructuring of the file upon input in this type of processing are directly affected by the form(s) of output desired. There are no outside factors, such as related file structure changes and/or interactions with which to be concerned. Therefore, the simplest form of file management software can be applied. As an illustration, the union list as now being provided consists of five output formats:

1. Main Entry - Union Listing
2. Author Main Entry
3. Subject Entry
4. Added Entry
5. Main Entry - Shelf Listing

The input records are formatted for key sort for each type of entry (main, subject, holding library, etc.), and then sorted by that key. Once in sequence, the keyed file is then accessed one record at a time and, depending upon the type of entry, the proper information is extracted from the original record to format the output record. All of the parameters used in determining the keys and formats are included within the controlling program. Any desired changes which would affect the output or input records would require a great deal of re-programming.

By using file management support software, only the key structure and record structure of the output files need be changed - not the program itself. The entire output is created by one pass of the input file instead of three as was necessary without the support of a file management processor (See Figure 12). Processing would be reduced to a fraction of the time required without such supporting software.

Batch (On-Line) Processing - Single File/Multiple Files

If the capability for on-line processing of the bibliographic data base is desired, even in the simplest form of retrospective cataloging (see Page 40 of this addendum), a change must be made in the type of file management support software used. At this juncture in the system development, it would be of great significiance to know what part the bibliographic data base will play in the overall function of the system. Will it always be used only to update holdings of the various user libraries and produce a union listing, individual user listings, regional listings or combinations of these; or will it eventually take the form of a true on-line catalog of holdings for individual users and for the entire State with file query and retrieval capability? In Appendix D of the study, it was pointed out that the cost and effectiveness of the various commercial data base management systems can vary

PROJECT/PROGRAM NAME	PROJECT NO.	PAGE OF
FILE MANAGEMENT - BATCH MODE		
CUSTOMER	PROGRAM NO.	DATE

FLOW CHART

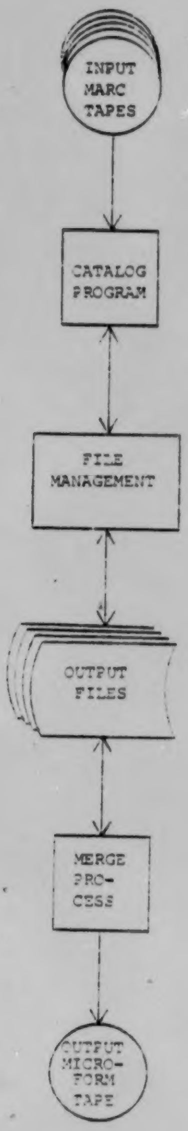

INPUT MARC TAPES
CATALOG PROGRAM
FILE MANAGEMENT
OUTPUT FILES
MERGE PROCESS
OUTPUT MICRO-FORM TAPE

FIGURE 12

PREPARED BY

immensely. If the future use of the data base is not to be in the form of on-line files and the interaction of these with the various other modules within the system (acquisitions, reference searching, inter-library loan, etc.) (Figure 2), then a data base management system such as IBM's IMS, Honeywell's IDS/11, Univac's DMS/90 or 1100 and even some of those offered by the manufacturers of so-called "mini" systems (Data General, Digital Equipment Corporation or Hewlett-Packard) might suffice. However, if the intent of the system is to eventually use the bibliographic data as the basis for all potential activities of the system, a more powerful and flexible data base management system should be incorporated into the network. Page D.4 of the study lists, in detail, the functions a data base management system should perform in order to effectively process a large relational data base such as a fully integrated library system. It was determined from the study of the various commercially available systems that only two were viable contenders: (Appendix D of study)

1. Software AG of North America's "ADABAS"
2. Cincom System Corporation's "TOTAL"

and even these do not provide all of the needed functions. Also, since they are generalized systems, many functions not needed to manipulate such a data base are included in the system and must be paid for, not only in actual dollar outlay for the purchase of the software, but in core and storage space utilization and in processing time. Alluded to but not pursued in the study was the finding that the most cost-effective solution might be to design and implement a data base management system as well as an associated query language which would fulfill all the desired functions.

Support software has not been written, at least in a commercially available form, to handle bibliographic MARC records, nor any kind of so-called "intellectual - relational" system based on the MARC format. If such a system for data handling and file query were written to incorporate all the functions listed on page D.4 of the study, the software system would also be able to process "natural language" systems if and when this technology becomes available. This is most important to any intellectual community such as a library since it will surely replace the MARC record format. MARC format is a stop-gap method for processing large bibliographic data bases in the same manner that microform catalogs are a stop-gap method employed between card and on-line catalogs which will eventually be available not only in the library but through a home TV screen.

NOTE: For further reading on this subject, the following is recommended:

1. An English language question answering system for a large relational data base; David L. Waltz; Communications of the ACM, July, 1978, Vol. 21, No. 7 (Attached at the end of this addendum).
2. DEACON (Direct English Access and Control system) G.E. Tempo - Rome Airforce Base.
3. Those references checked at the end of Article 1.

Query Language Software: There are perhaps six prime features that must be present in any query language software:

1. The language of implementation must be transportable
2. The language of implementation must be independent of the operational language
3. It must provide accessability to all data base files - regardless of the inherent file structure
4. The operators must provide normal compares and full Boolean logic

5. The ability to report selected information to the user must be provided

6. File sorts, tallys and counts

It is not necessary to have a data management system in order to process query language software. The only relation between the two lies in the organization of the data file structures provided by data management software. If the data base management system provides good file structure and linkage paths, then the query language software need not do so and the entire process of selection from the data base becomes faster and more efficient.

Language of implementation: The programming language chosen for the implementation of the query software must be transportable; i.e., the programs must be able to be processed on any computer which supports the programming language chosen. The languages of implementation must be oriented toward system programming techniques (such as PASCAL, ALGOL, PL/1) and not one oriented toward applications or number crunching (such as COBOL, FORTRAN, BASIC). This is most important within the context of an integrated library network. All application modules must be able to query and draw conclusions from the various files which make up the total data base. The application modules will likely be processed on a mix of computers. If the query software is transportable, it can be used regardless of the hardware involved.

Independence from operational language: The operational language is the programming environment used to write the various application modules and other system support software. As noted in the section on data base management systems (See page D.3 study) the implementation software for the data base management system is such that any application programming language which allows for CALL's to the data management system software may be used (such as COBOL, FORTRAN, ASSEMBLER, PL/1, etc.). In other words, each is independent of the other. This same relationship should exist between the query language software and other operational software used within the system.

This independence from all operational software is lacking in the commercially available query languages such as ADABAS' ADASCRIBT, IBM's IQF, IDS/11's Data Query System, etc. Independence does exist between the application software and the query language software but NOT between the data base management and query language software. Therefore, in order to process any of these query systems, the associated data management software must be present within the system.

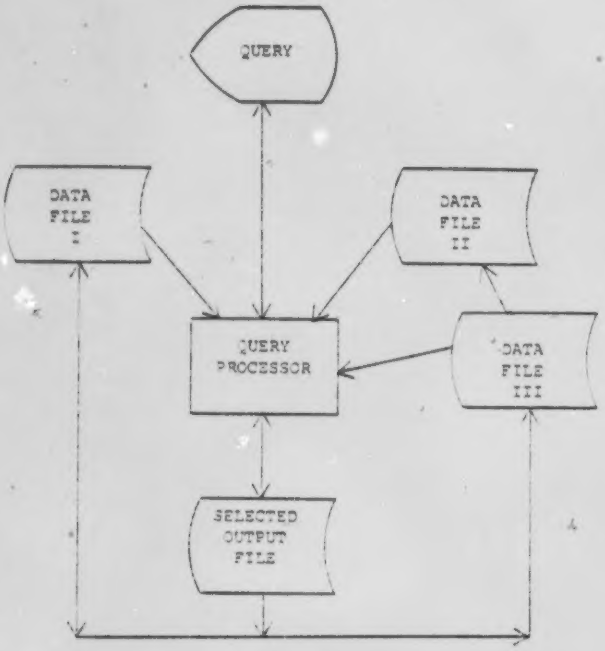
The technology needed to program a completely independent query language is available and a few software concerns are attempting to provide such a system commercially. (The Computer Company is now in the process of implementing such a system.) The process involves a "marriage" of known data processing, natural language processing and data base management processing techniques.

Data Base Accessibility: (Figure 13). The query software must have the capability to:

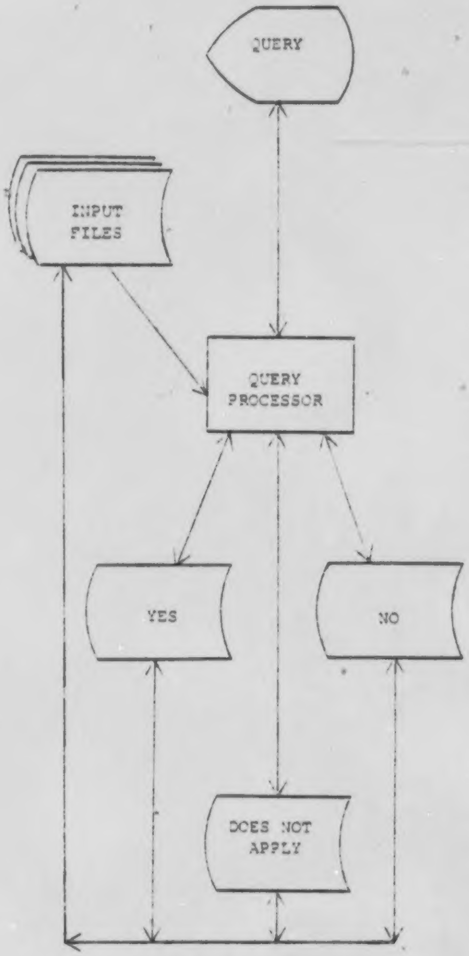
1. From one select string, gather information from all files within the data base, as opposed to simply one file at a time
2. Gather information from any field within any record in a file
3. Create intermediary files of selected records which in turn can be queried individually or in combination with other files
4. Partition or separate a file or files as determined from the select string

PROJECT/PROGRAM NAME DATA BASE ACCESSABILITY	PROJECT NO.	PAGE OF
CUSTOMER	PROGRAM NO.	DATE

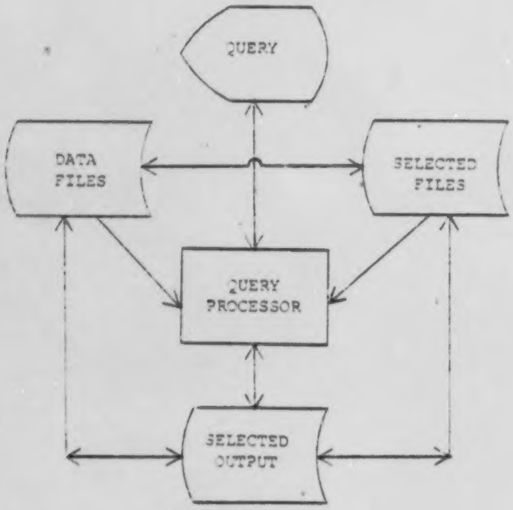
FLOW CHART


QUERY
DATA FILE I
DATA FILE II
QUERY PROCESSOR
DATA FILE III
SELECTED OUTPUT FILE

(1) MULTIPLE FILE ACCESS


QUERY
INPUT FILES
QUERY PROCESSOR
YES
NO
DOES NOT APPLY

(3) FILE PARTITION


QUERY
DATA FILES
SELECTED FILES
QUERY PROCESSOR
SELECTED OUTPUT

(2) COMBINATION - DATA BASE AND INTERMEDIARY SELECTED FILES

FIGURE 13

THE COMPUTER COMPANY - T961 (11/74)

PREPARED BY ______

5. Retrieve by:
    a. Compare (direct)
    b. Compare (partial)
    c. Key
    d. Dictionary look-up
    e. Inversion
    f. Imbedded key
    g. Imbedded compare

Multiple Access: In order to query multiple files, the query language system must be capable of processing Boolean arithmetic, (Logical and, or, etc.). The use of this technique in conjunction with the accessibility of all files within the data base allows the system to provide the user with logical comparative answers to queries which range across the entire data base. (See Operators and Boolean Logic Page 67).

This makes it possible to obtain information from any file within the data base regardless of its structure. Page 4 gives an illustration of the relational data structure needed to programmatically obtain all requested information from one access to the data base. This same information can be obtained through the use of query software alone.

Intermediary Files: The query language system must be able to collect into an intermediary file all records answering the selection string criteria. The intermediary file should contain pointers back to the original record(s) as opposed to a copy of the record itself. By structuring in this manner, while the intermediary file exists, access to it or the use of it in any manner will ensure that the information is current. These files must become a logical part of the data base structure so that they may be queried in the same manner as the original data file (Figure 13(2)) and be used in a "select down" mode. Therefore, if from one select string 100 records are retrieved, the user may now more finitely select from this intermediary file of 100 records.

Partition Files: The query software must allow for the partitioning of a file. (Figure 13(3)). To illustrate: A user library wishes to obtain three listings of information from the purchase order file:

a. List of fiction monographs on order
b. List of non-fiction monographs on order
c. All others, regardless of type, on order

An example of the input for such a query follows:

IN PO, Type EQ M, F-IX; Type EQ F, F-2X;
Type NEQ M or F, F3X.

This string would translate as follows:

In the purchase order file do the following -

1. All records whose type is equal to M add to file 1X
2. All records whose type is equal to F add to file 2X
3. Add all others to file 3X

NOTE: The original purchase order file has not been changed. Each of the intermediary files contains the addresses only of the record within the original file for quick retrieval of the record for print, display or any other use desired.

Retrieval Criteria: The problem of retrieval from a bibliographic file has been discussed numerous times, both in the study and in this addendum. To be effective throughout the entire system the query language must provide at least 7 forms of retrieval. (Figure 14)

RETRIEVAL

TYPE	MEANING	EXAMPLE
DIRECT COMPARE	THE INPUT STRING IS COMPARED EXACTLY - BYTE FOR BYTE	INPUT: AUTHOR EQ HEMINGWAY ALL RECORDS WITH HEMINGWAY IN THE AUTHOR FIELD WILL BE RETRIEVED
PARTIAL COMPARE	ONLY PART OF INPUT STRING IS COMPARE	INPUT: P AUTHOR EQ JOHN.. TON ALL RECORDS WITH THE FIRST FOUR CHARACTERS OF THE AUTHOR FIELD EQ TO 'JOHN' AND LAST THREE CHARACTERS EQ TO 'TON' WILL BE RETRIEVED
KEY	CERTAIN PREDEFINED NUMBER OF CHARACTERS IN PREDEFINED COMBINATIONS ARE GIVEN AS INPUT STRING	INPUT: AUTTIT HEMI & MOO ALL RECORDS WHOSE FIRST FOUR CHARACTERS IN THE AUTHOR FIELD 'HEMI' AND FIRST THREE LETTERS IN TITLE FIELD ARE 'MOO' WILL BE RETRIEVED

FIGURE 14

RETRIEVAL

TYPE	MEANING	ILLUSTRATION
DICTIONARY LOOK UP	A DICTIONARY OR DIRECTORY OF LIKE TERMS IS MAINTAINED BY THE SYSTEM WITH LINKAGE PATHS TO WHERE IN THE DATA BASE THESE TERMS MAY BE FOUND	INPUT: U.S. DEPT. AGRICULTURE ALL RECORDS CONTAINING: U.S. DEPT. AGRICULTURE UNITED STATES DEPT. AGRICULTURE AGRICULTURE U.S. DEPT. OF OR ANY OTHER PRE-DEFINED FORM OF THE STRING WILL BE RETRIEVED
INVERSION	COMPARISON MADE ON INVERTED FIELD	INPUT: U.S. DEPT. AGRICULTURE ALL RECORDS CONTAINING: U.S. DEPT. AGRICULTURE and AGRICULTURE U.S. DEPT. WILL BE RETRIEVED
IMBEDDED KEY	PRE-DEFINED KEY STRUCTURE IS FORMED FROM CHARACTERS WITHIN THE FIELD	INPUT: USCOMECO ALL RECORDS WHICH HAVE: US COM ECO

TYPE	MEANING	ILLUSTRATION
		AT BEGINNING OF WORD WOULD BE RETRIEVED U.S. COMMERCE DEPT., ECONOMICS
IMBEDDED COMPARE	THE INPUT STRING IS SOMEWHERE WITHIN THE FIELD	INPUT: ECONOMICS ALL RECORDS WITH THE STRING 'ECONOMICS' IN THE FIELD WILL BE RETRIEVED U.S. DEPT. AGRICULTURE, ECONOMICS ECONOMICS - FORECASTING ETC.

These forms must be applicable to all fields and sub-fields, as opposed to only certain fields (Author, Author/Title; ISBN, etc.) as is the case in most bibliographic data bases.

Operators: The system must provide a full set of operators which determine how the data is to be selected from the data base:

Equal to
Not Equal to
Less than
Greater than
Less than or equal
Greater than or equal
Between
Outside of

and combinations of these in conjunction with the Boolean operators:

and	any
or	none
	if
	then

Reporting: The capability to create output formats on-line is desireable (Report Generation) but not necessary as long as pre-defined formats are available and will accept a selected (intermediary) file as input.

Sort/Counts: The system must allow for sorting of the file using any field within the record structure of the file. It must also provide the capability of obtaining totals of any numeric field within the record and of counting the number of records present in the selected file.

Once the selection has begun the system must allow the user to stop the process at any time, reporting to the user:

Number of records looked at
Number selected
Number not selected

The system should then allow the user to either continue or terminate the process.

Communication Software: The amount and complexity of communication support software cannot be defined until the network actually exists. However, it can be stated that any communication hardware and/or software must be programmatically addressable so that changes may be made to the network, especially in protocol, as needed (see section C of study).

Operating System: The underlying system software must be able to support:

1. All types of file structure:
    a. Sequential
    b. Index Sequential
    c. Multi-indexed
    d. Etc.
2. Variable length records and variable length fields within records.
3. Multiple files
4. Broad range of programming languages
    a. COBOL
    b. BASIC
    c. FORTRAN
    d. PL/1
    e. PASCAL
    f. ASSEMBLER
    g. Etc.

5. Various Modes of Operation:
    - a. BATCH (On-Line or Off-Line)
    - b. Real-Time/On-Line
    - c. Mix of Modes

6. Multi-type terminals and multi-use of terminals

Integration Requirements: The bibliographic data base file must be accessible through any application module supported by the system. It must, to be efficient, be relational to all other data files within the system.

Available Systems: Section 3 and B of the study report in some depth on three major commercially available bibliographic services and the projected use of such services within the context of an integrated library system.

Pages 29-34 of this addendum report on what is now being done in the State and what further action could be undertaken to create a bibliographic data base for Virginia.

Time and Cost Projections: Cataloging Module

It is neither the intent nor the purpose of this addendum to recommend specific brands of hardware or support software nor to provide exact hardware configurations for a proposed network. The study makes it very clear that many items of information needed to produce a meaningful network configuration are simply not available at this time. Those involved in the study have gathered quite a bit of insight as to how the network should evolve (Section 3 of Study). Therefore, this section and all succeeding sections on Time and Cost Projections will be based on the following assumptions:

1. The initial design of the system will take into consideration both short and long term use of the system and provide for these from the beginning.
2. The system will be used in some way by the majority of the library community.
3. The evolution of the system will be modular (not necessarily one at a time) and enough time will be allowed to design, implement, and integrate each module effectively within the system.
4. The evolving system will be autonomous in the sense that all system components, especially processors and storage devices, are dedicated to the system and not shared with other data processing users. The size and complexity of a fully integrated on-line system makes this absolutely necessary.

Hardware: Cataloging

Processor

First and most important, any processor decided upon must be upgradeable without affecting any of the application or support software of the system. It must, from the beginning, have enough main memory to support:

1. Bibliographic data base
2. Program development and testing
3. Data Base Management system
4. Query language system
5. Communications, real-time and batch
6. Cataloging module programs

It must also have the following features:

1. Maximum 200 Nanosecond cycle time
2. VS (Virtual Storage) capability
3. Be upgradable to 2,000,000 bytes of main memory and 4,000,000,000 bytes of mass on-line storage

The purchase price for such a processor ranges from $500,000 to $1,240,000 depending upon the amount of memory required. Monthly leasing ranges from $12,600 to $32,900. Monthly maintenance costs range from $1,150 to $2,600.

Mass Storage

The on-line mass storage device must contain at least 1,000,000,000 bytes and must be expandable to 4,000,000,000 bytes. Purchase prices and leasing costs for mass storage devices which are compatible with all commercial mainframes vary greatly. Any price given here would be meaningless.

Terminals

The cataloging module presents certain problems in terminal selection both for data entry display (CRT) and for hard copy printing. The ideal situation when designing an integrated system would be the use of 'universal' terminals throughout the network, terminals which are compatible with commercial communication protocol and which also could be used regardless of the application being processed by the system.

On-line cataloging, in its truest form, seems to require for data entry and display a terminal capable of accepting foreign language characters and diacritics, or at least, the ability to accept and translate character combinations to achieve this. The display of these foreign characters and diacritics must also be taken into consideration. Further study should be done to determine if, through a combination of hardware and software techniques, it would be more cost-effective to apply these techniques to a universal terminal than to require specially designed terminals. The cost of a terminal seems insignificant, at first glance, in comparison with other hardware components, but this cost must be multiplied by several hundred and maybe thousands depending upon the number of users of the network. The cost ranges from less than $1,000.00 per unit to over $5,000.00 per unit, which is quite a spread.

However, a good universal non-intelligent CRT may be purchased for approximately $1,200.00 per unit and intelligent terminals (allowing some memory and storage within the unit itself) may be purchased for under $3,000.00. On-line cataloging units could easily be of the non-intelligent group.

The choice of a hard copy terminal (printer) is definitely dependent upon the functions the terminal will perform within the system. The type of printer needed for each application module could vary from a small strip printer attached to a CRT for issuing receipts, etc., to a large, high speed printer to produce card cataloging. The price differential between such printers is great, as is the price range within each type. The most important facts to keep in mind during design is that:

1. All printers must use a standard EIA board
2. If catalog card production is to be a feature offered by the system, the printer(s) chosen must be capable of handling full ALA character sets.

Communication Hardware/Line Charges

The kind and complexity of communication hardware can only be determined when some idea of the number of users of the system is known. This applies to line charges also. What grades of communication lines will be needed? What would be the most cost-effective line configuration?

250

However, there are certain features which must be provided by any communication hardware considered:

1. The capability to communicate with a mix of processors and/or terminals
2. Be programmable
3. Automatic or programmable switching mechanism between processors and/or terminals
4. The capability to be upgraded and/or expanded as the need arises
5. The capability now, or at least in a development stage, to use satellite and/or radio wave as medium of communication as opposed to hard wire.

The cost of this hardware in the beginning is dependent upon the configuration of the network. It can range from a few hundred dollars (if used only to connect terminals to a central site for processing only one or two application programs) to several thousands of dollars (if used to communicate and manage the communications between several satellite processors and/or terminals).

Line costs are best considered during the actual network topological design phase. A typical network line configuration is shown in Figure 19. The total number of line miles is 249. This figure times the cost per mile for the grade of line desired would give line charges per month.

The on-line cataloging module for retrospective cataloging would be done in the beginning using normal dial-up telephone lines. This is quite inexpensive. However, plans should include the upgrading of these lines to 9600 BAUD and direct leased line capability before actual on-line cataloging, data base search or any other application is added to the system.

Application Software:

Time and cost projections for application software to process the cataloging module is dependent upon such factors as:

1. The familiarity of the analysts, designers and programmers with the cataloging process and with its inherent problems.
2. What the scope of the process will be:
    a. Retrospective cataloging only
    b. On-line direct cataloging
    c. Reference searching
    d. Integration with other processes and/or data files within the system
    e. Authority files for control added to the system

  f. Programming language restrictions by the user. (Must it be written in COBOL or may other programming languages be used?)

  g. The type of support software available to the programmer (data base management system, operating system, etc.).

3. Will the design, analysis and programming groups be dedicated to the project?

4. Will the machine used for development, testing, and implementation be dedicated to the project?

These are not known. Therefore, time and cost projections for the cataloging module will be based on the following assumptions:

1. The team involved has some knowledge of the data base content, the overall system objectives, and problems that have been encountered by others in the design and implementation of such a module.
2. Good support software is available for use, mainly in the form of a data base management system, operating system and communication system.
3. Hardware capacity allows for efficient programming, testing, and file handling.
4. Team involved in design, analysis and programming is dedicated to the project.

Even with these assumptions, the time given for design, implementation and testing are purely ballpark.

1. On-line retrospective cataloging
   8 man-months analysis and design
   4 man-months programming
   1 man-month testing/implementation

2. On-line cataloging - no authority control
   15 man-months analysis and design
   27 man-months programming
   4 man-months testing/implementation

3. Addition of authority control
   5 man-months analysis and design
   15 man-months programming
   2 man-months testing/implementation

4. Addition of LC MARC files to on-line cataloging module
   5 man-months analysis and design
   10 man-months programming
   2 man-months testing/implementation

The enitre cataloging module from inception (allowing for retrospective cataloging only) to full use (including all sub-files within the cataloging module and "hooks" for integration with other planned modules):

Analysis and design	33 Man-Months
Programming	56 Man-Months
Testing/Implementation	10 Man-Months

Lead time for hardware and supporting system software is dependent upon the facilities the State, or a vendor chosen by the State, may already have in place. To start from scratch, at least 6-12 months should be allowed for the securing, testing, and implementation of the hardware and support software. (See page 126 for further statistics)

Topological Configuration:

The actual placement of hardware for the cataloging module and all subsequent modules depends upon the number of users and their geographical density. For purposes of illustration, we will assume that Richmond will be the site of the central bibliographic data base. In actuality, this may or may not be the case. The placement of hardware must be flexible. Any network that is built around the premise that a certain location, and certain equipment resident within that location, must be used as a central or distributed node of the network would defeat the purpose of complete independence of hardware location to the internal workings of the network.

Figure 15 is an illustration of the network as it might look for on-line retrospective cataloging.

Figure 16 illustrates the addition of on-line cataloging.

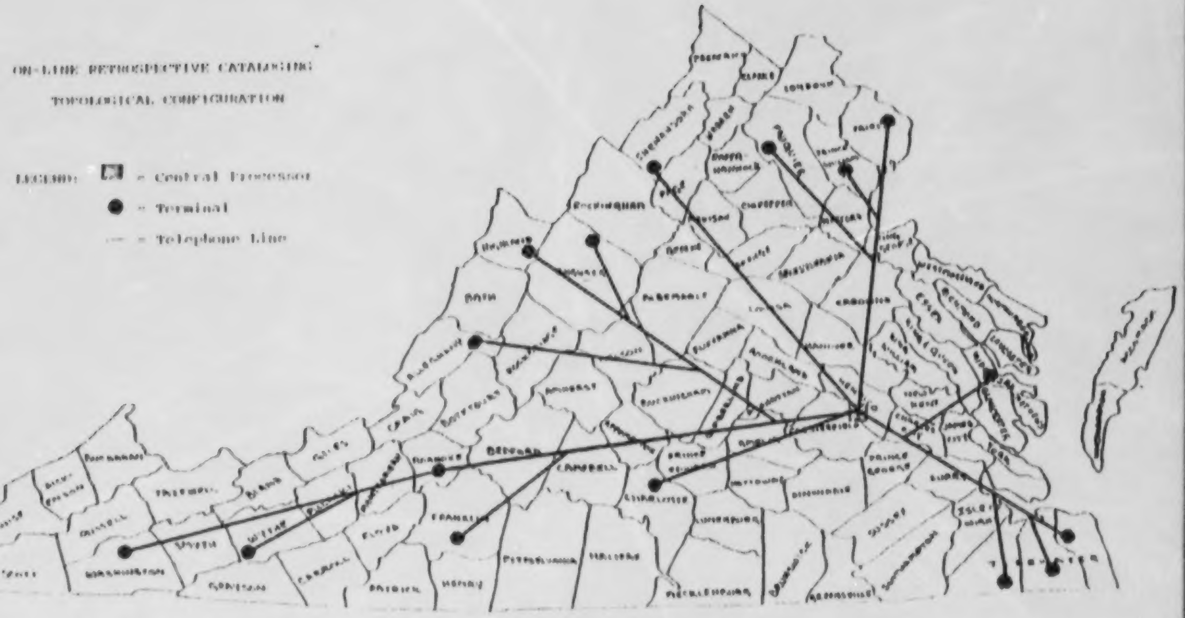
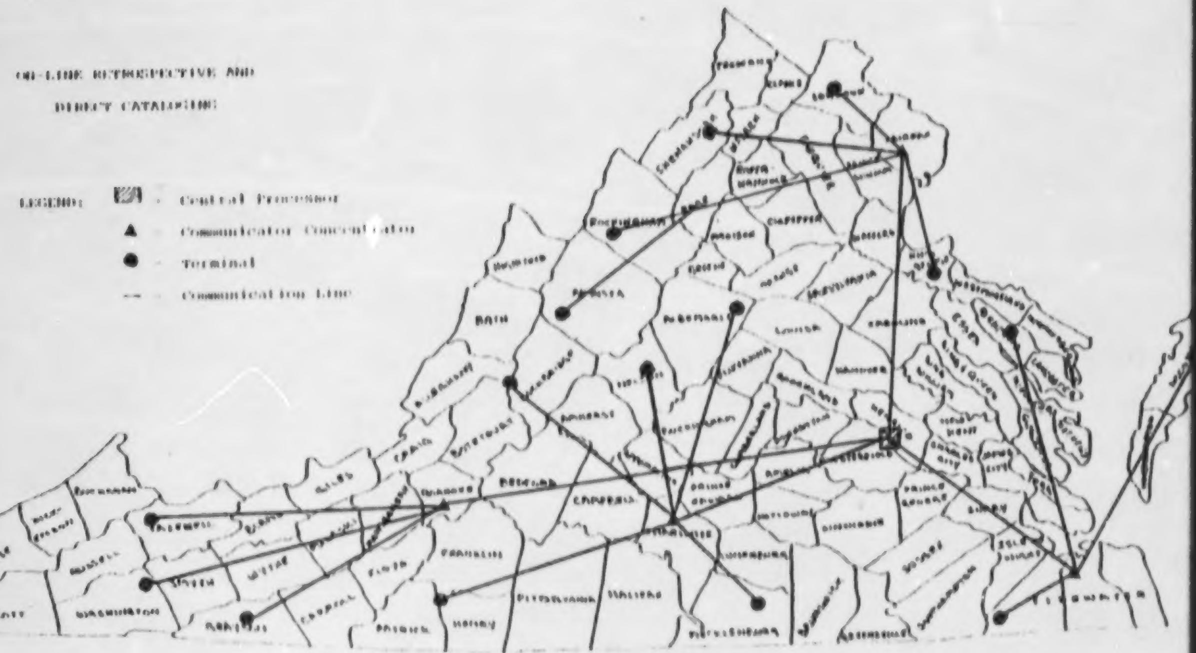

ON-LINE RETROSPECTIVE CATALOGING
TOPOLOGICAL CONFIGURATION
LEGEND:
▣ = Central Processor
● = Terminal
— = Telephone Line
BATH
ROCKINGHAM
ALBEMARLE
LOUDOUN
CLARKE
GILES
BEDFORD
CAMPBELL
FRANKLIN
FLOYD
SMYTH
WASHINGTON
LEE
SCOTT
PATRICK
HENRY
PITTSYLVANIA
HALIFAX
MECKLENBURG
DINWIDDIE
SUSSEX

FIGURE 15

ON-LINE RETROSPECTIVE AND

DIRECT CATALOGING



LEGEND:
▨ - Central Processor
▲ - Communicator Concentrator
● - Terminal
— - Communication Line
LEE
SCOTT
WISE
RUSSELL
WASHINGTON
GILES
CRAIG
BATH
FRANKLIN
CAMPBELL
PITTSYLVANIA
HALIFAX
HENRY
PATRICK
LOUISA
CULPEPER
LUNENBURG

FIGURE 16

Figure 16 shows the addition of front-end miniprocessors to handle the data flow to the central site. These processors act only as communication links at this time, with some storage capability. Their purpose is to accumulate information until the central processor is ready to receive it for processing or, in case of failure or overload of one or more nodes in the network, to switch the data to other processors in order to keep a balance within the system.

Communication linkage between the terminals and the processors could consist of a mix of direct and dial-up lines. Communication to the central processor should be direct high-speed lines.

Conclusions and Recommendations:

Cataloging Application Module

1. The data base of MARC records as it currently exists (OCLC retrospective tapes), should be expanded to include other machine readable records available in the State.
2. File management capability should be added to the processing of the file to produce catalogs in microform.
3. The catalogs should be updated more frequently than once every six months as is now planned.
4. Serious consideration should be given the various forms, other than commercial vendors, of retrospective cataloging, especially to include the use of the existing data base.
5. The bibliographic file(s) should be considered the center of the eventual system.
6. Once on line, the file structure must be relational and not relative.
7. All application programs must be able to function regardless of the mode of operation.
8. The system must have full backup and security controls.
9. The form and content of any output from the system, for each individual user or group of users, must be easily changeable.
10. Hardware requirements are dependent upon the use of the system. However, any hardware must be state of the art and must be easily upgraded and/or expanded as the need occurs.
11. The ability to efficiently manage and query the system is of utmost importance. An in-depth analysis should be made regarding commercial data base systems and query languages, and a comparison should be made between the.

use of one of these and the design and implementation of such systems by the State or by a software firm designated by the State.

12. The data base should include:

    a. Virginia base of MARC records

    b. LC MARC records (monographs and serials)

    c. Holding records

    d. Authority files

Circulation:

Section 3 of the study recommends that the circulation application module be planned and designed at the same time as the on-line portion of the cataloging module. During a preliminary analysis (as was conducted during the study) of the software components necessary to create a fully integrated library networking system, one usually finds one or two application modules which if available would answer 80 percent of the need for such a system. The cataloging and circulation modules of a library network provide this since the data contained in the bibliographic and circulation files are used directly to process:

1. On-line and retrospective cataloging
2. Circulation and finding
3. Reserve
4. Interlibrary loan
5. Catalog searching and control

and indirectly to process:

1. Acquisitions
2. Serials check in

3. Etc.

Figure 16.1 outlines the relationship that must exist between these bases. The actual physical location of the bases is not important so long as the hardware and supporting software for the system allow this relationship to exist. (See prior sections on data base and cataloging).

Data Base Content

The circulation module should contain at least three major files: patron, item (book, film, serials, etc.), and transactions. These files must be completely relational to one another as well as to all other major files within the data base.

Patron File

The record content of the patron file should have at least the following:

1. Name
2. Address
   City
   State
   Zip Code

PROJECT/PROGRAM NAME CATALOG & CIRCULATION MODULE DATA RELATIONSHIP		PROJECT NO.	PAGE OF
CUSTOMER		PROGRAM NO.	DATE

FLOW CHART

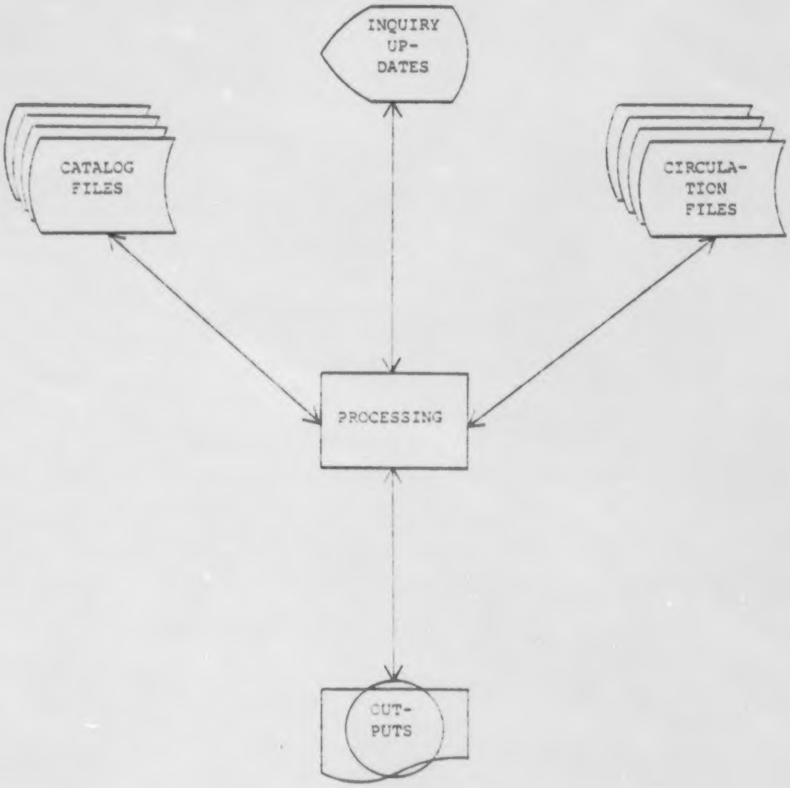

INQUIRY UP-DATES
CATALOG FILES
CIRCULA-TION FILES
PROCESSING
OUT-PUTS

FIGURE 16.1

PREPARED BY ____________

3. Phone number
4. Card number
5. Type of borrower code (adult, child, senior citizen, student, faculty, etc.)
6. Status code (past due, etc.)
7. Number of volumes checked out at present
8. Number of volumes check out in past
9. Date last transaction

The minimum size of such a record would be approximately 123 bytes. Using the latest available statistics (1976 - 1977), there are 2,101,860 registered patrons within the state. (This does not include private colleges, schools, and special libraries). Storage requirements for this file would be 123 x 2,101,860 or 250,210,780 bytes of storage for the patron file alone. Coupling this with statistics given for the number of circulation transactions in one year (26,803,711), it becomes quite apparent that the circulation module would probably be more effective if distributed to satellite processors rather than it being resident in a central processor thus giving faster accessability to the data. Further investigation of these statistics makes it evident that the State can be divided into five geographic regions (Figure 17):

1. Capital Area
2. Northern Virginia
3. Tidewater
4. Piedmont
5. Western

Even regionalized, the amount of storage for this single file is quite large.

Region	Registered Patrons	File Size
Capital	346,885	42,666,855 bytes
Northern	641,144	78,860,712 bytes
Tidewater	520,941	64,075,743 bytes
Piedmont	120,808	14,859,384 bytes
Western	472,082	58,066,086 bytes

The volatile, non-static nature of a library data base, which was discussed earlier in this addendum, now becomes even more significant. Even if one considers a seven percent growth rate, the file will eventually become quite unwieldly. In the prior section on cataloging, the importance of allowing for variable length records and variable length fields within records was stressed repeatedly.

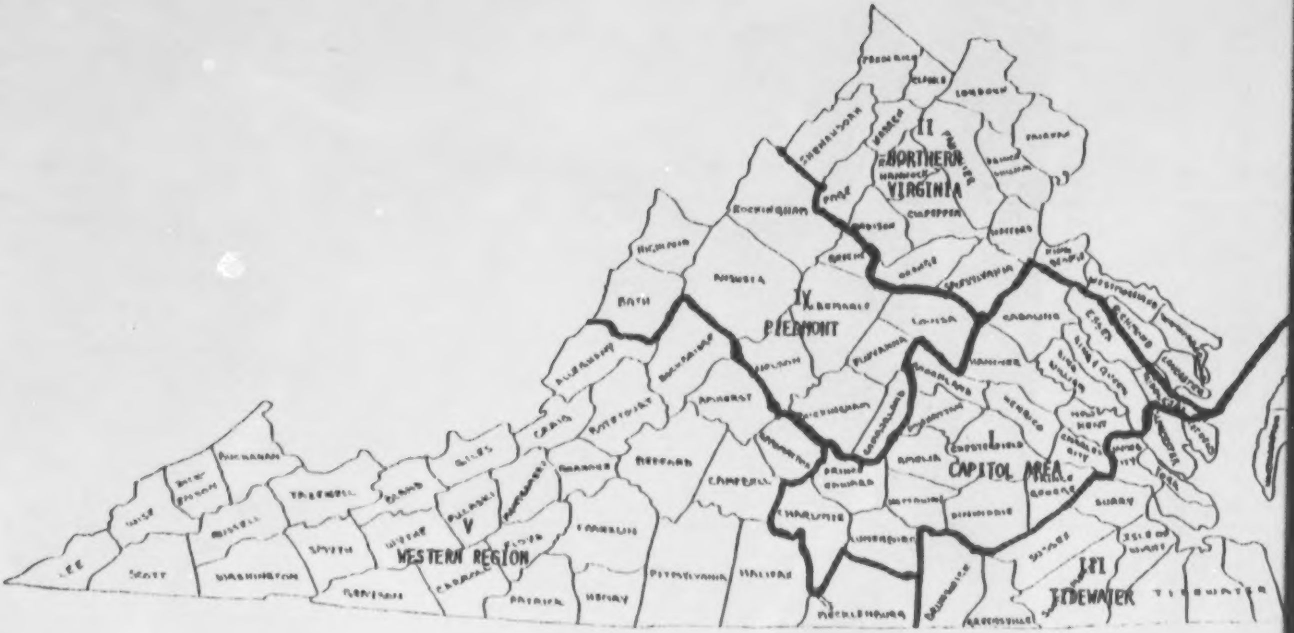

II
NORTHERN
VIRGINIA
IV
PIEDMONT
I
CAPITOL AREA
III
TIDEWATER
TIDEWATER
V
WESTERN REGION
LEE
SCOTT
WISE
BUCHANAN
RUSSELL
TAZEWELL
WASHINGTON
SMYTH
WYTHE
BLAND
GILES
PULASKI
CARROLL
GRAYSON
PATRICK
HENRY
FRANKLIN
CRAIG
BATH
HIGHLAND
AUGUSTA
ROCKINGHAM
AMHERST
BEDFORD
CAMPBELL
PITTSYLVANIA
HALIFAX
MECKLENBURG
CHARLOTTE
LOUISA
HANOVER
HENRICO
AMELIA
DINWIDDIE
SURRY
CAROLINE
ESSEX
FAIRFAX
LOUDOUN
CLARKE
FREDERICK
CULPEPPER
SHENANDOAH

FIGURE 17

In dealing with a data base which is volatile and non-static, the handling of the storage and retrieval of the data in an efficient manner is probably the most important design criteria.

Second in importance is to eliminate as much duplicate information as possible. The patron file, therefore, must be structured in such a way that it will allow for one entry of the demographic information (address, telephone number, etc.) per household with the other household members' records acting as 'trailers' of this master:

Example:

MASTER RECORD

FAMILY NO.
MEMBER NO.
NAME
ADDRESS
PHONE
TYPE OF BORROWER
STATUS
NO. VOLS PRESENTLY OUT
NO. VOLS PAST
DATE LAST TRANS
POINTER (CIRCULATION)
POINTER (TRAILER)

↔

TRAILER RECORD

FAMILY NO.
MEMBER NO.
NAME
TYPE
STATUS
ETC

Through the use of this file structure, we have eliminated the duplication of address and telephone number data. If the design of the data base is relational and the proper data base management system is being used (see pages 23, 50), then it matters little where the master record is located.

This concept of file and sub-file relationship can be carried even further when discussing the next file within the circulation data base.

Item File

The item or circulation record can be represented in at least two different forms, dependent upon the presence of the title in the bibliographic data base.

As mentioned earlier, duplication of data within the data base must be kept to an absolute minimum. In theory, a relational data base has no data repeated, but in fact it is sometimes necessary to duplicate portions of the data in order to achieve the response time required by a given application. This may be true for circulation and for the bibliographic information contained in the circulation record. (i.e., Title, Author, etc.) A truly integrated library system contains one glaring difference from most other systems - the sharing of one large base of information which is common to all (bibliographic) and from which most of the information for the other bases (circulation, acquisitions, etc.) may be extracted. Whether the total data base can be managed well enough so that the bibliographic information can be extracted from the bibliographic base for each function within a given module cannot be determined until the scope of the network is known.

The design of the system should address this condition and allow for changes in file content and data base structure to accommodate adjustments which must be made in order to achieve the response time needed for any application process within the system.

The circulation record content could, therefore, be in one of two forms:

A. Desired bibliographic information obtained from bibliographic file:

MASTER RECORD

KEY TO BIBLIOGRAPHIC FILE
(The information retrieved decided by individual user)

ANY STATISTICS THAT USER MAY DESIRE FOR MANAGEMENT PURPOSES

TRAILER RECORD - ONE FOR EACH COPY OF TITLE

LABEL NUMBER
LIBRARY DESIGNATION
BRANCH DESIGNATION
CLASSIFICATION NUMBER
COPY NUMBER
PRICE
SOURCE
STATES FLAG
POINTER TO MASTER RECORD

Using this type of structure means that each master record would contain a minimum of 20 bytes. From the statistics available, we find that there are a total of 9,018,836 titles of which 10% are probably unique. (This does not include private schools, public schools or special libraries). Then 901,884 x 20 would equal the storage needed for the master records (18,037,672 bytes).

Each trailer record would contain a minimum of 50 bytes. Duplicate copies of all titles number 6,333,414. 50 x 6,333,414 would equal the amount of storage needed for trailer records (316,670,700 bytes) or a total of 335,708,372 bytes of storage for the entire file.

Regionalized, the statistics are:

270

	MASTER RECORD (bytes)	TRAILER RECORD (bytes)	TOTAL (bytes)
Capital	273,390	47,218,300	47,491,690
Northern	2,976,700	78,328,050	81,304,750
Tidewater	4,385,246	64,758,150	69,143,396
Piedmont	2,926,824	66,046,100	68,972,924
Western	5,014,996	60,320,600	65,335,596

Serials are not included in these figures because of the lack of available statistics. Once gathered, they should be added to the total size.

B. Bibliographic information contained in the circulation record:

This would not appreciably increase the size of each trailer record but would definitely increase the total storage size. Even if a minimum of 50 bytes (which may not be realistic) were allowed for main entry, title, etc., the storage needed would increase:

	FROM (bytes)	TO (bytes)
State Wide	316,670,700	379,807,580
Capital	47,491,670	56,786,950
Northern	81,304,750	88,749,930
Tidewater	69,143,396	80,106,490
Piedmont	68,927,924	76,289,970
Western	65,335,596	77,873,100

a rather significant increase especially if potential growth factors are applied.

Transactions File

This file, unlike the patron and item file, would likely be cleared periodically. Its function within the module is to act as backup in the event of system failure, either the entire system or a regional processor. However, the potential size of this file must be considered, as storage space must be provided. The minimum size of each record would probably be 80 bytes. To obtain some idea of the amount of storage required, we will assume that the file will be purged daily and that daily circulation for the entire State is approximately 26,803,711 volumes per year. (From 1976-77 statistics.) If we assume 300 working days in a year, then the daily circulation would be approximately 87,345. This amount times the record length per transaction (80) equals the amount of storage that must be available for the transaction file (6,987,600 bytes), if the process is centralized.

By Region: (distributed processing)

REGION	APPX DAILY CIRCULATION	STORAGE
Capital	13,191	1,055,280 bytes
Northern	32,037	2,562,960 bytes
Tidewater	18,496	1,479,680 bytes
Piedmont	7,382	590,560 bytes
Western	18,234	1,458,720 bytes

By totaling the storage needed for each file (Patron, Item, Transaction), we can determine the total amount of storage needed for the circulation module.

	TOTAL FILE STORAGE
Capital	91,223,825
Northern	162,728,422
Tidewater	134,698,819
Piedmont	84,422,868
Western	124,860,402
State Wide	594,906,752

These numbers are based on 1976-77 statistics and do not include private colleges, public schools or special libraries. Normal growth rate and the addition of private colleges, public schools and special libraries will increase the file size. Also, storage must be allotted for application software, system support software, and file overhead, which will normally be another 25-30%.

Input Methods

The module must allow for on-line/real-time input as well as for batch input (in case of system failure).

Processing:

Program Requirements

1. On-line file maintenance programs for the patron and circulation file (pages 45-46)
2. Process check-in and check-out transactions
3. Capability to allow user to change loan periods, by type of borrower and/or by item type
4. Must calculate from date checked-out - date due, taking into account holidays and days the library is not open. This must be easily changed by the user.
5. Provide easy access to individual title or group of like titles
6. Ability to read and translate machine-readable labels
7. Ability to update all files from one transaction
8. Take reserves and update relevant files, regardless of their location within the total system
9. Provide for automatic inter-library loans - locating material, updating proper files and creating output for managerial and informational purposes throughout the entire system. (See page 118 Inter-library loans)
10. Handle fine accounting by automatically calculating fines and/or charges and printing overdue notices

11. Alert the user if patron has overdue items, fines due, too many items out, during check-out. Allow for user change of parameters and for override with security control.

12. Handle serials.

File Query - Security Control

Those items outlined in the Cataloging/Bibliographic Module should apply to the Circulation Module.

Form and Content of Output

1. Printed overdue notices
2. Printed reserve notices
3. Printed receipts (if desired)
4. Management reporting

Hardware Requirements

Any processor chosen must be able to accommodate the growth of files, the addition of terminals and the addition of mass storage devices with no degradation in the specified response time of 2-5 seconds. The processor must be large enough to handle any portion of the data base management system, query system, communication system and operating system which must be core resident. The minimum size of a processor (for a region) should be 512KB of memory, expandable to 2MB of memory. The processor should have the capability to off-load all I/O processing to alleviate any response time degregation which might occur as terminals, storage devices, etc., are added to the system.

Communication

See page 75 cataloging module.

Peripherals

The data terminals (CRTS) must have some type of tape backup capability in case of system failure. The terminal must be able to detect the failure and initiate the backup procedure. It must also detect when the system is again functional and begin the process of updating the files from the transactions gathered during system down time.

Terminals must also collect transactions and process these in a batch mode at a later time.

Printers must use a standard EIA board and the system configuration must allow for upgrading the printer terminal at any time.

Mass storage (disks) should be easily added, with 200MB being the minimum amount of storage, but with ability to add as needed.

Software Support Requirements

See page 50 Cataloging Module.

System Integration Requirements

All files which comprise the circulation data base, regardless of their location in the system, must be related to all other files in the system. (See pages 4-5 file structures).

Available Systems

Section F of the Study reports on some of the commercial turn-key systems available. Section 3 describes how these system might be integrated into the projected network, some of their good points and some of their drawbacks. Many pages of this addendum have been devoted to the importance of a total design concept and the use of a relational file structure. To incorporate any of these commercial systems into this design concept would be virtually impossible. There is, however, available in the State an on-line cataloging and finding system which addresses all of the circulation design criteria and could be modified to adhere to the overall relational structure of the proposed integrated system (VPI system). The application module is written in COBOL and, therefore, is transportable to most well-known hardware (mainframe or mini). The system has been in operation long enough that those involved in the design and implementation and those using the system are in a position to know not only what is good about the system, but of greater help, what should be changed and how it should be changed. Certainly working from such a base would be more cost-effective than beginning from scratch. It also might be more cost-effective, in the long run, than purchasing turn-key systems and giving up the space-saving, time-saving efficiency that is acquired from a relational data structure. A cost savings could be realized if the hardware used within the network is purchased or leased directly from the manufacturer than if purchased from a vendor.

Time and Cost Projections

Time and cost projections for the hardware and supporting software needed to process the circulation module at a centralized site is included in the section on Cataloging. Time and cost projections for mini systems can range from under $5,000.00 to over $1,500,000.00 depending upon the amount of memory, storage and number of terminals needed. To put the problem of costing in better perspective, there follows a price comparison between a quoted figure for a given configuration of a turn-key system (Data Phase) and the price of the system as taken from the hardware manufacturer's price list. The prices quoted from the book are list prices and do not include discounts, or any special considerations which could easily decrease the cost of each component by 15-30%.

The configuration given by the vendor was priced at from $200,000.00 to $250,000.00. It is not known if the difference of $50,000.00 lies in changes to software or the addition of terminals over and above the four which are supplied with the system.

	CONFIGURATION	MANUFACTURER LIST PRICE
(1)	ECLIPSE S/230 COMPUTER WITH 64KB MAIN MEMORY	$19,000.00

(1)	192MB DISC STORAGE SYSTEM WITH CONTROLLER AND ADAPTER FOR UP TO 4 DRIVES	$33,000.00
(1)	MAGNETIC TAPE DRIVE	9,900.00
(1)	PRINTER	3,200.00
(16)	DISPLAY TERMINALS	33,600.00
	TOTAL	$103,300.00

for communication modems and multiplexors added:

	5,000.00
TOTAL HARDWARE LIST PRICE	$108,300.00

This then means that the vendor is charging $141,700.00 for any system support software and application software used by the system. This amount will remain constant regardless of the size of the processor or the number of storage units purchased.

Application Software

It was just determined what value one commercial vendor puts on its application software. It is impossible, without a complete investigation of the programs in the circulation system available in the

State and a knowledge of what hardware and supporting software will be available, to determine the cost of enhancing the State system to meet the criteria for a relational system. A good guess would be that $150,000.00 be adequate.

Support Software

See cataloging module - page 50

Topological Configurations

Figures 19 and 20 are examples of both a distributed and a centralized system respectively.

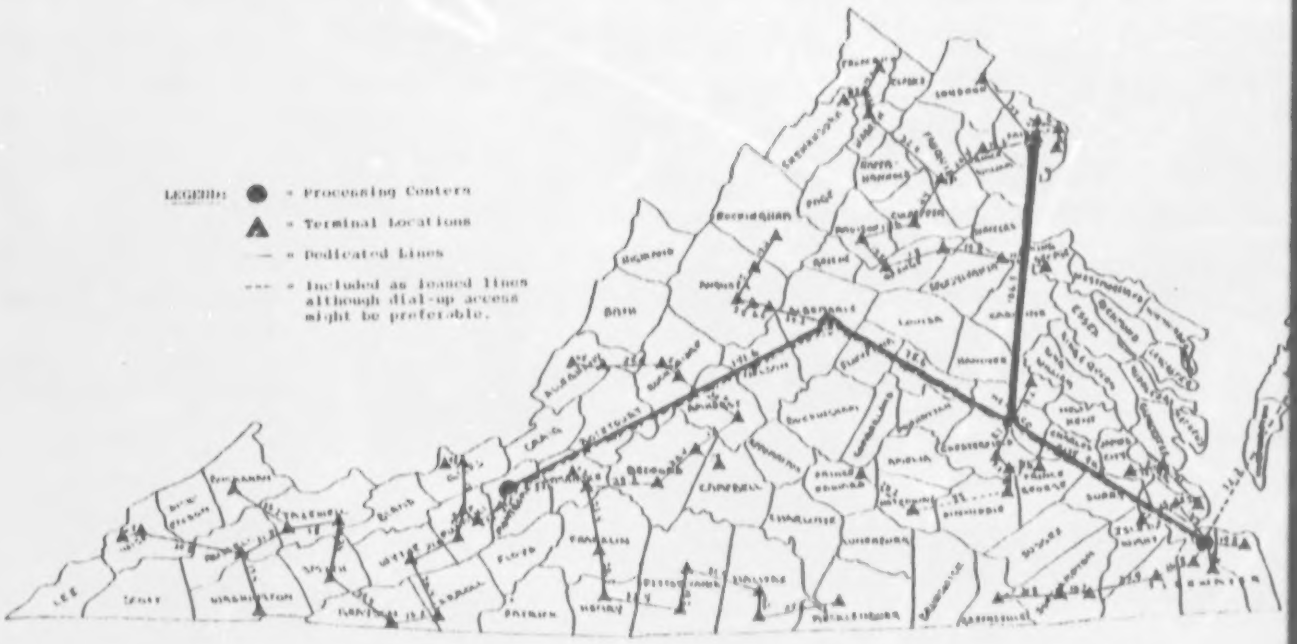

LEGEND: ● = Processing Centers
▲ = Terminal Locations
— = Dedicated Lines
--- = Included as leased lines although dial-up access might be preferable.
LEE
SCOTT
WASHINGTON
FLOYD
FRANKLIN
PATRICK
HENRY
CRAIG
BATH
BEDFORD
AMHERST
CAMPBELL
CHARLOTTE
LOUISA
HANOVER
AMELIA
LOUDOUN
CULPEPER

FIGURE 19

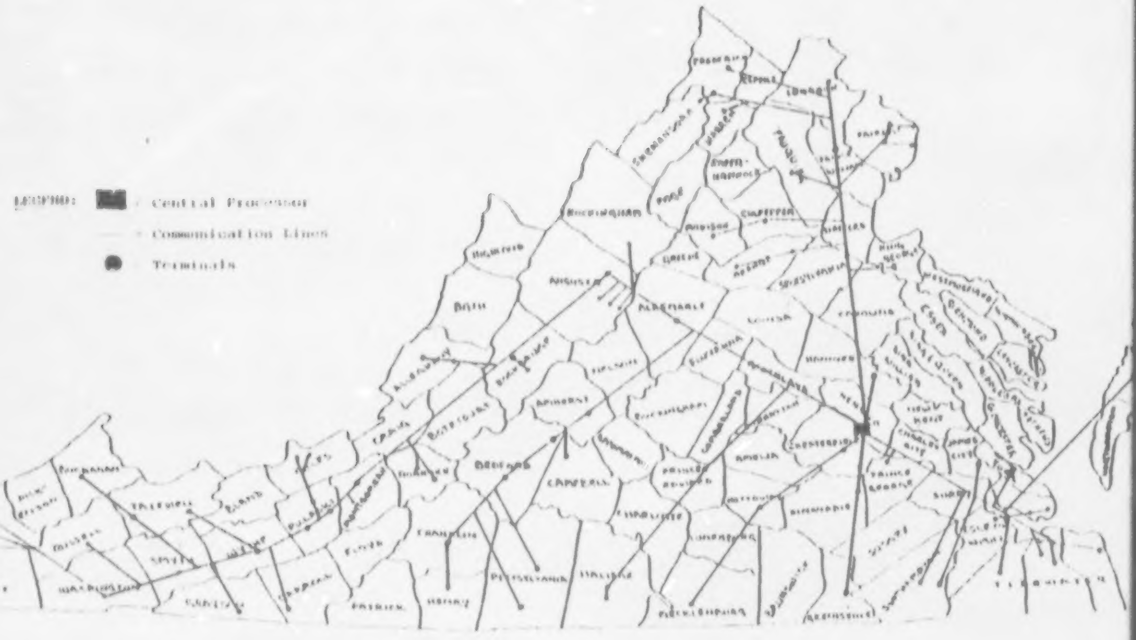

LEGEND: ■ - Central Processor
— - Communication Lines
● - Terminals
LEE
SCOTT
WASHINGTON
BUCHANAN
DICKENSON
RUSSELL
TAZEWELL
BLAND
GILES
SMYTH
WYTHE
PULASKI
MONTGOMERY
CARROLL
FLOYD
PATRICK
HENRY
FRANKLIN
PITTSYLVANIA
HALIFAX
BEDFORD
CAMPBELL
AMHERST
NELSON
ALBEMARLE
AUGUSTA
BATH
HIGHLAND
ROCKINGHAM
ALLEGHANY
BOTETOURT
ROCKBRIDGE
CRAIG
ROANOKE
FLUVANNA
BUCKINGHAM
CUMBERLAND
PRINCE EDWARD
CHARLOTTE
LUNENBURG
MECKLENBURG
BRUNSWICK
NOTTOWAY
DINWIDDIE
AMELIA
CHESTERFIELD
HENRICO
HANOVER
LOUISA
CAROLINE
SPOTSYLVANIA
ORANGE
GREENE
MADISON
CULPEPER
RAPPAHANNOCK
PAGE
SHENANDOAH
FREDERICK
CLARKE
LOUDOUN
FAUQUIER
FAIRFAX
STAFFORD
KING GEORGE
WESTMORELAND
CHARLES CITY
JAMES CITY
PRINCE GEORGE
SURRY
SUSSEX
SOUTHAMPTON
GREENSVILLE

FIGURE 20

Acquisition Module

The data base files contained in the acquisition module are probably the most controversial, in terms of accessibility by any member in the network. Therefore, page 19 of this addendum discusses the need to protect certain files from global use. The manner in which this data is protected is a decision that must be made during the analysis and design stages of the system. However, two examples will be given here:

1. When a purchase order is entered into the system the following would occur:

    a. Circulation record created with status bit set to 'on order'

    b. Purchase order file updated

    c. Accounting file updated.

    Using this method, any user of the network would know simply by the normal search of the circulation file that the title is to be available and where it will be available. The actual updating of the bibliographic files should not be done until the user informs the system, through the acquisition module, that the volume(s) have been received.

2. The acquisition module could have an 'on order' file. The content of this file would be:

    a. A pointer to the title in the bibliographic file

    b. The number of copies on order

    c. Purchasing library code

    d. Pointer to P.O. file.

As in example one, any further action should update all relevant files and reset the data pointers as needed. Again, all users would have access to volume on order without having access to actual accounting data.

Data Base Content

The entire thrust of this addendum concerning the evolution of an integrated library system has been toward data base purity coupled with the use of relational file structure. Example 1 on the previous page meets these criteria better than 2. Therefore, the acquisition data base, as defined here, will not have an 'on order' file. There are five basic files that would normally comprise the acquisition data base:

1. Vendor file
2. Purchase order file
3. Invoice file
4. Accounting file
5. Transaction history file.

Vendor file content:

1. Vendor control number
2. Vendor name
3. Vendor address

4. Type code
5. Pointer to transaction file
6. Account control information

The approximate size of each record depends upon the amount of control information held in the record. It may range from 150-300 bytes.

Purchase order file content

Master Record

1. Pointer to vendor file
2. Pointer to library control file
3. Purchase order number
4. Status code
5. Shipping code
6. Date
7. Any other information that is needed and is common to the entire order

Trailer Record - one for each unique item on the Invoice

1. Pointer to bibliographic or circulation file
2. Pointer to master record
3. Vendor bock number
4. Number ordered
5. Number received

6. Actual cost
7. Status code
8. Responsibility code
9. Any other information needed and unique to this item.

Approximate size of Master record: 50 bytes.

Approximate size of each trailer record: 70 bytes.

Transactional History File Content

Two types of transactions will be created.

1. Purchase order
2. Invoice

The content of the purchase order transaction record should contain all the information needed to update all files related to the issuing of the purchase order:

1. Circulation
2. Vendor
3. Accounting
4. Purchase order
5. Transaction

The content of the invoice transaction record should contain all the information needed to update all files related to the receipt of an invoice:

1. Bibliographic
2. Circulation
3. Purchase order
4. Accounting
5. Vendor
6. Transaction

Accounting file

The content of the accounting record(s) within the file is dependent upon the depth of information desired by each user library. Further analysis must be done before an actual layout can be presented. From past experience, however, 200 bytes per record should be adequate. The file structure should coincide with the accounting procedures used by the individual user (chart of accounts). Estimated file sizes will not be given for the acquisition module as was for cataloging and circulation because of the lack of statistics. However, the following information may be useful during the analysis and design of the system:

1. The purchase order file will be relatively static, since it should be purged when all items have been received.

2. The vendor file will also be more or less static. If control accounting is kept in the vendor file, then the file structure must be:

    a. A vendor file for each accounting group, or

    b. A master vendor record common to all, with trailer records for each accounting group.

3. Accounting file: Also more or less static. However, it must be structured in such a way that additions or deletions of master, sub or sub-sub fields can be easily accomplished by the user.

4. The following statistics are known about the number of items acquired. These are statistics for 1976-77:

	Volumes	Titles	Periodicals
Entire State:	987,034	476,031	77,768
Capital Area:	166,575	75,499	12,280
Northern Area:	255,279	86,285	9,191
Tidewater Area:	214,885	117,128	15,979
Piedmont Area:	105,758	62,311	18,486
Western Area:	244,537	134,808	21,832

Fig. 21 illustrates the acquisition module data base relationship.

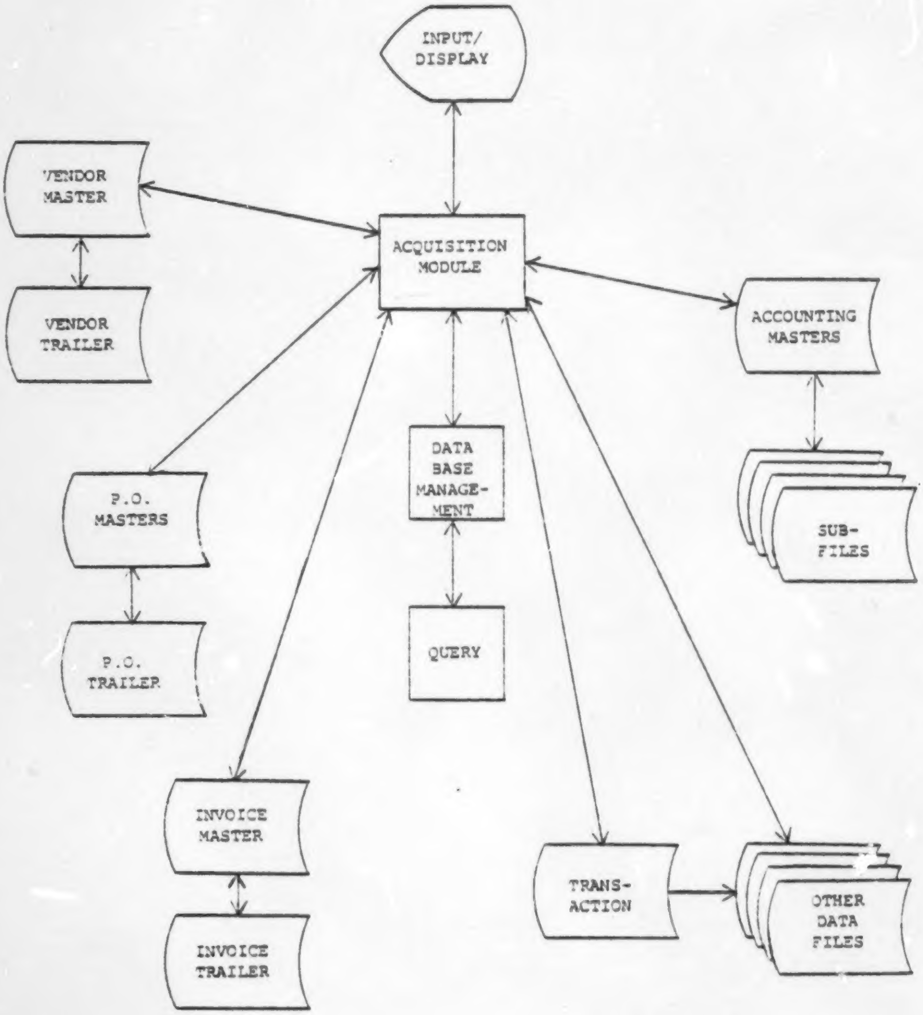
Input Methods

Although accessability to the data base files (for inquiry, creation of purchase orders and matching of invoices) must be tolerable to the user, it is less critical here than in the circulation module.

PROJECT/PROGRAM NAME ACQUISITION MODULE FILE	PROJECT NO	PAGE OF	
CUSTOMER STRUCTURE	PROGRAM NO.	DATE	

FLOW CHART

A B C D E F G H I J K L M N O



INPUT/ DISPLAY
VENDOR MASTER
ACQUISITION MODULE
VENDOR TRAILER
ACCOUNTING MASTERS
DATA BASE MANAGE- MENT
P.O. MASTERS
SUB- FILES
QUERY
P.O. TRAILER
INVOICE MASTER
TRANS- ACTION
OTHER DATA FILES
INVOICE TRAILER

FIGURE 21

A B C D E F G H I J K L M N O

THE COMPUTER COMPANY – T961 (11/74)

PREPARED BY: ____________________

Also, the use of the acquisition module will probably be quite varied. The module must allow for all modes of operations:

1. Batch (off-line)
2. Batch (on-line)
3. On-line/real-time
4. Combinations of the above

Processing Acquisitions:

Program Requirements:

The entire process of circulation and acquisition can be likened to "accounts payable and inventory control," the main difference being that items are simply added to the inventory and very rarely deleted. Acquisitions is the accounts payable module of such a system and will probably contain more individual programs than any other module in the system.

1. File maintenance programs:

   Allow the user to add, change and delete records contained in the vendor file, purchase order file, invoice file, accounting file.

2. Data Input Program:

   a. Allow the user to input data concerning vendors, purchase orders, invoices, payments, cancellations, back-ordering, price changes, debit and credit adjustments, etc.

b. Allow user to display and/or print any record or group of records from any file.

c. Print purchase orders, invoices, checks, ledgers and any other form deemed necessary by the user.

d. Managerial and statistical reporting on demand.

5. Purging Programs:

Allow user to purge transactional history file and invoice file on demand. The program must produce adequate audit trails and file updating in conjunction with the purge.

Software Support Requirements

Are the same as outlined in the cataloging and circulation modules.

System Integration Requirements

All files within the acquisition module must have access to each other as well as to every other file within the total data base. All files within the base must be updated as needed from any transaction created by the acquisition module.

Available Systems

In section 3 of the study we find that very little is commercially available and those which are available must be purchased in conjunction with a turn-key system. To write the module from scratch would probably take at least 6 man-months for analysis and design, 36 man-months to program and 3-4 man-months to implement and test. However, there are within the State two fairly sophisticated on-line systems. The programming language used for each is transportable and allows for calls to most data base management systems. These two systems should be investigated in detail as a potential base for the acquisition module. Changes would be required in order to have the files work within a relational structure but, as with the in-state circulation system, the benefits could be extremely cost-effective.

File Query, Data Control, Backup Measures

See pages 56, 19.

Hardware Requirements

Hardware requirements for the acquisition module are similar to the circulation module. In a total system, with regional satellite processors, the acquisition module would reside in the same location as the circulation module for the user library. How much more storage and processor space must be allowed for the addition of this module cannot be determined without further analysis. Since most of the files within the module are static (with the exception of the transactional history and invoice file) and not large in size, additional storage is not a primary concern. The transactional history and invoice files will be periodically purged, so these files will also remain static. Consideration should be given to the size and type of processor, since the addition of several terminals could degrade response time throughout the system. If the processor chosen has a programmable IOP (input-output processor) to handle all terminals, the degradation can be easily eliminated. If not, this feature would need to be added, which might mean a complete change in processor. An alternative solution would be to add more main storage, if possible. If neither alternative is feasible, then a new processor must be installed. This again points up the concern, stressed so often in the addendum, that an overall analysis and design must be done before proceding too far in determining the hardware components for the network.

Inter-library Loan Module

Unlike the modules discussed so far, the inter-library loan module does not have any data files related directly to it. The module uses files throughout the system:

1. Bibliographic
2. Circulation
3. Acquisition

Also, unlike the other modules it does not have a master control program which allows a choice by the user of various procedures to be processed. To be truly effective, the programs which make up the inter-library loan module should be callable as subroutines from other modules within the system:

1. Circulation
2. Reference

By so doing, the entire inter-library loan process can be initiated from the terminal being used for circulation or reference searching with very little effort.

The importance of good communication hardware and software has been addressed with each module of the system. The inter-library loan module is probably most dependent upon this capability. To be effective, the entire process of search and reserve message sending should be transparent to the user. Also the system must know if the holding library for the volume desired will allow the automatic reservation of the item or if a message of request must be issued. The analysis and design of the inter-library loan module should clearly depict the method by which the satellite files will be searched, at what times during the search the user will be expected to make decisions pertaining to the continuance or termination of the search and in what sequence the various files will be accessed. The communication hardware and software should be flexible enough to alter normal routing if a processor is not functioning or if the line load is too heavy. The user should be made aware if certain files are not accessed due to system failure. Some type of time interval should be set by the user before each search, thus enabling the search to be discontinued if desired. The module must also be able to:

1. Update all applicable files
2. Create messages for display and/or print
3. Provide a complete audit trail of transactions
4. Produce daily and on-demand reports as desired by the user.

Available Systems

There are no on-line systems available at this time that fit the network concept.

Time and Cost Projections

The amount of time needed to program the inter-library loan module is dependent upon:

1. The data base management system being used by the network
2. The communication hardware and software support being used by the network
3. The power and sophistication of the query language software being used by the network.

If the three support systems listed above have the power and sophistication recommended in the first section of this addendum, the inter-library loan module consists of little more than a control program whose function is to gather parameters (external and internal), provide message displays and gather and format hard-copy output. Within this context, the following times are realistic:

1. Analysis and design: 3 man-months
2. Programming: 3 man-months
3. Testing and implementation: 4 man-months

On the other hand, if sophisticated system support utilities are not available, the programming of the module would be useless. The time required to find the item within the data base itself would turn most users back to the telephone request system.

Serials Check-in and Control - Reference Searching

There are two other functions which must be performed by the system. As with the inter-library loan function, these have no data files but act upon the files of other modules in the system.

Serials are handled through three modules:

1. Acquisitions
2. Cataloging
3. Circulation

The design of each module should provide for serials acquisition, cataloging, circulation and check-in. When these three modules are implemented, the automatic control of serials will be accomplished.

Reference Searching

This is handled by the data base mangement system, communication system and query language software.

Network Approaches

As determined by the group concerned with the Study, there are many approaches which can be taken in the evolution of a library network. However, there are, among the many, three approaches which seem reasonable:

1. Build the network from the bibliographic base of holdings
2. Build the network from regional bases
3. Combination of the above

The third of these is the recommended approach of the study group. However, all are viable alternatives and much the same results could be achieved by using any of these approaches.

The first sections of the addendum and section 3 of the Study report in detail the reasoning behind the recommendation of an overall conceptual approach to a network as opposed to a fragmented approach. The reasoning will not be repeated here. Instead, we shall outline each approach, with benchmarks to be achieved within each before the next step is taken.

With all three approaches the assumption is that the final goal is a total network encompassing all library functions.

Seven basic steps should be taken in order to create an integrated networking system:

1. Overall analysis of the system and of the functions it is to perform
2. Analysis of the potential users of the system
3. Overall design of the various functions and their interrelationships
4. Detailed design of the beginning module(s) of the system
5. Analysis of the hardware, supporting software, and application software needed to support the functions. Cost analysis of alternative configurations of hardware and support software.
6. Personnel to design, program, implement, test, and administer the evolution of the system and the network
7. Purchase or lease of the hardware and software components needed to begin the network.

This addendum has attempted to provide some analysis of the functions needed to achieve an integrated library system, but certainly not in the depth required to attempt to begin the design.

Bibliographic Approach

Under this approach the State would continue building, using the Solinet archival tapes as the beginning central base of holdings and adding other holdings.

This would be accomplished through the use of computer-aided retrospective conversion of the existing card catalogs and available machine readable records. A mix of modes (batch, on-line) would be used for this process. (See page 41 addendum.) Once this module is implemented and tested, the next module will begin.

Considering the seven basic steps we find that:

1. Overall analysis: This addendum does supply the information needed to determine what relationships will exist between the bibliographic (cataloging) module and future components of the system.
2. An analysis must be done to determine the potential users and/or suppliers of the module.
3. An overall design of the system must be done.
4. Detailed design of the cataloging module, its functions, data structure, file structure and data content must be done.
5. Analysis of hardware and supporting software.

The hardware needed to accomplish this could be either a mainframe (370/145, etc.) or a series of large capacity minicomputers (which are, in reality, mainframes) such as Data General M600 series, DEC 20 series or TANDEM.

Most importantly, the hardware must be dedicated to the processing of the library system.

The support software must be able to provide the tools needed for the cataloging module and for the integration of future modules. The cataloging module itself demands good support software for data base management and query language.

This same sophistication should be applied to the communication hardware and software. Its need may be less critical now than in the future, but to change it later could be more costly in time and money than to supply it at the beginning.

6. Steps 4, 5, and 6 really must envolve in tandem. Those involved in the design and implementation of the system must also be involved in the hardware, support software, application programming medium and overall path the system will take. It is of utmost importance that members of this team be dedicated to the project and answer only to those administering the project.

7. Once the hardware and support software are purchased, integrated and tested programming can begin. In the analysis of the cataloging module, it was determined that lead time to secure the hardware would be at least 6-12 months. The analysis, design, programming, testing and implementation of the cataloging module alone would take approximately 8.25 man years. If we assume normal commercial rates for design and programming the cost for this phase could be close to $500,000.00 over a probable two year period. Hardware and supporting software costs could range from $500,000.00 to $2,000,000.00. Outright purchase of the hardware would not be the best approach. Leasing agreements, with no penalty for upgrades and expansion, are available and are more cost-effective.

The initial overall system analysis and design should take approximately two man-years or one calendar year, at a commercial cost of approximately $150,000.00.

Regional Bases

Using this approach the State would be divided into geographical regions and the data bases on each regional processor would be accessible from any other processor within the network. This approach would differ from the centralized approach since regional processing would more likely begin with a circulation module, with other modules (excluding cataloging) added as required. The assumption here is that the circulation records can now be converted to full MARC records and eventually used to create an on-line common bibliographic file, accessible by all regions.

Steps 1 through 3 of the seven basic steps toward the development of an integrated system must be completed before attempting step 4. From the first three steps it can be determined whether to purchase vendor-supplied turn-key systems and depend upon the vendor to supply the necessary future modules. In addition, determine if these systems will fit into the design concept of

1. Relational data files
2. Data base management system geared to the handling of intelligent data
3. Powerful query language capable of decision making

or to use as a base the circulation and acquisition systems available in the state, enhance these and provide homogenous hardware and support software for them.

If the former is chosen then steps 4 through 7 will be performed by the selected turn-key vendor or vendors.

If, however, the latter is chosen, step 4 now addresses itself to analysis and design of the changes that will be necessary in the existing systems to allow for complete file relationship, interaction with the data base management system, operating system, and any other supporting software and/or hardware.

Step 5 - hardware analysis should still occur even though the circulation and acquisition systems alluded to are both using the same hardware (Hewlett-Packard). It may be found that this equipment fits all the criteria of the proposed system and its cost is no more than for other manufacturers. However, the opposite might also be true. It is likely that much of this analysis will have been done during steps 1 through 3. A sample configuration, using data from this addendum, might be useful to compare turn-key quoted costs with in-state development.

The Northern Virginia region can be used as an example. We will assume for the purpose of this example that the system must support:

1. 60 locations with 2 terminals per location (120) (60 of intelligent type, 60 non-intelligent).
2. Total mass storage needed at present is 200,000,000 bytes and another 200,000,000 bytes should be provided for future growth, support software, application software, etc.
3. Processor size must be 512KB to support all processing.
4. Programmable communication hardware at processor site and modems and/or multiplexor at terminal sites.
5. All hard-copy printing is to be done at processor site and two 600 LPM printers are needed.
6. 1600 BPI tape drive unit is needed for backup and other purposes.
7. An IOP (input/output processor) is needed to take terminal load off the main processor.

The hardware list price would be approximately $552,000. The cost of modifying the existing circulation system approximately $150,000 (however, this is a one-time cost).

If the data base management system and query language system are designed and implemented in-house, the cost could be $250,000-$350,000. (Again this is a one-time cost.) This is reasonable in comparison with the cost of commercial systems (see section D of study).

This amounts to nearly $1,000,000 for the first region. Subsequent regions would require only the hardware expenditure. Also, the data base management and query system could be used with all additional modules, including cataloging.

It is difficult to estimate the time involved to accomplish this. An estimate (assuming a strong development group) would be one to two calendar years to become fully operational.

This may seem long in comparison with the installation of a turn-key system, but the benefits to be derived from a total design concept likely outweigh the time required for development.

Combination Central and Regional

Section 3 of the study describes in detail a modular approach to an integrated library system, allowing for parallel development of a centralized base of common holdings and satellite bases for processing circulation, acquisition, and other related functions. The addendum to the study describes in detail how each of these sub-systems (functions) must interrelate and what components (hardware and software) are needed to achieve a totally integrated system.

The combined approach is recommended by members of the study group. There are many reasons for selecting this approach over the others. Using the same seven steps:

1. - 3. The overall analysis of the system functions, potential users and overall design should be the same regardless of the approach.

4. The detailed design of the beginning modules now takes on a whole new scope. The relational file concept will not only be designed and implemented, but now tested at the inception of the system.
5. Analysis of hardware for both central and distributed bases will be done at the same time. This will lead to a better hardware and topological configuration of the system.
6. More efficient use of the personnel involved in the project. The time and cost savings would be quite significant.
7. Purchasing or leasing hardware from one or two vendors can decrease costs by 20-30%.

By combining the two approaches and taking advantage of the cost savings realized by more efficient use of personnel, both the centralized bibliographic module and the distributed circulation module (including the supporting data base management and query systems) could be accomplished within two or three calendar years at a cost saving of about 20% or $180,000. Savings on the equipment lease or purchase again would be 20-30%.

Lastly, as was stated in the study, both systems are needed in order to address the needs of the majority of the library community now and in the near future.

Administration

The problem of network administration has only been alluded to in this addendum. The manner in which the net work is governed is not the concern of those preparing this study, but we feel we must make some recommendations as to the administration of the network:

1. The hardware used must be dedicated to the sole function of processing the library system.
2. The personnel involved in analysis, design, programming, implementation and testing must be autonomous and answer only to those governing the network.

3. It should be the function of this group to administer the data processing functions of the network.

    a. Programming updates/changes

    b. Data base management

    c. File updating

    d. Analysis reporting

    e. Output distribution

    f. Management liaison

    g. New requirements analysis, programming, installation and testing

    h. Facility management and trouble-shooting

    i. User training

and be directly answerable to those governing the network.